

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 11:34:37 ; Search time 12.23 Seconds
(Without alignments)
13,469 Million cell updates/sec

Title: US-09-589-777A-2_COPY_168_175
Perfect score: 41
Sequence: 1 SYIVLCIE 8

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/prodata/2/iaa/5A.COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5B.COMB.pep:*
3: /cgn2_6/prodata/2/iaa/6A.COMB.pep:*
4: /cgn2_6/prodata/2/iaa/6B.COMB.pep:*
5: /cgn2_6/prodata/2/iaa/PCTUS.COMB.pep:*
6: /cgn2_6/prodata/2/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	185	3	US-08-985-526-36
2	41	100.0	195	1	US-08-159-784-2
3	38	92.7	183	4	US-09-206-059-2
4	30	73.2	154	3	US-09-191-647-10
5	30	73.2	191	1	US-08-159-784-3
6	30	73.2	254	1	US-08-948-616-9
7	30	73.2	254	4	US-09-193-510-9
8	30	73.2	254	4	US-09-368-402-9
9	30	73.2	307	2	US-08-948-616-3
10	30	73.2	307	2	US-09-193-510-3
11	30	73.2	307	4	US-09-368-402-3
12	30	73.2	395	1	US-08-723-938-3
13	30	73.2	395	1	US-09-080-538-3
14	30	73.2	420	4	US-09-008-271A-4
15	30	73.2	420	4	US-08-974-691-8
16	30	73.2	470	2	US-08-724-394A-10
17	30	73.2	775	2	US-08-714-070A-1
18	29	70.7	113	2	US-08-466-860-8
19	29	70.7	113	3	US-08-472-040A-8
20	29	70.7	113	4	US-08-276-776-8
21	29	70.7	113	4	US-08-471-209-8
22	29	70.7	343	2	US-08-878-989-5
23	29	70.7	343	4	US-09-272-796-5
24	29	70.7	1503	4	US-08-976-255-14
25	28	68.3	318	4	US-09-387-574-2
26	28	68.3	353	3	US-08-984-288-2
27	28	68.3	353	4	US-09-224-426-4

28	28	68.3	353	4	US-09-478-601-4	Sequence 4, Appl
29	28	68.3	400	5	PCT-US95-16472-2	Sequence 2, Appl
30	28	68.3	402	3	US-08-602-809-2	Sequence 2, Appl
31	28	68.3	422	4	US-09-224-426-2	Sequence 2, Appl
32	28	68.3	422	4	US-09-478-601-2	Sequence 2, Appl
33	28	68.3	468	4	US-09-355-115-8	Sequence 8, Appl
34	28	68.3	516	2	US-09-019-201A-3	Sequence 3, Appl
35	28	68.3	751	3	US-08-946-026-6	Sequence 6, Appl
36	28	68.3	4544	1	US-08-469-486-52	Sequence 52, Appl
37	28	68.3	4544	2	US-08-469-658-52	Sequence 52, Appl
38	27	65.9	22	2	US-09-013-634-8	Sequence 8, Appl
39	27	65.9	337	2	US-09-013-634-2	Sequence 7, Appl
40	27	65.9	471	3	US-09-106-464-2	Sequence 2, Appl
41	27	65.9	956	2	US-08-897-443-3	Sequence 3, Appl
42	27	65.9	1090	3	US-08-307-896-3	Sequence 3, Appl
43	27	65.9	1090	3	US-08-726-214-4	Sequence 4, Appl
44	27	65.9	1090	5	PCT-US95-11808-3	Sequence 3, Appl
45	27	65.9	1235	1	US-08-118-101A-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-985-526-36
Sequence 16, Application US/08985526
Patent No. 6080728
GENERAL INFORMATION:
APPLICANT: Mixson, James A
TITLE OF INVENTION: CARRIER-DNA COMPLEXES CONTAINING DNA
TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
TITLE OF INVENTION: THERAPY
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Connolly, Bove, Lodge, & Hutz
STREET: 1220 Market Street, P.O. Box 2207
CITY: Wilmington
STATE: Delaware
COUNTRY: U.S.A.
ZIP: 19899
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,526
CLASSIFICATION:
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/608,845
FILING DATE: 16-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: McMorow Jr., Robert G
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 658-9141
TELEFAX: (302) 658-5613
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 185 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-985-526-36

Query Match 100.0%; Score 41; DB 3; Length 185;
Best Local Similarity 100.0%; Pred. No. 0.58;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 169 SYIVLCIE 176

RESULT 2
US-08-159-784-2
Sequence 2, Application US/08159784
Patent No. 5643783
GENERAL INFORMATION:
APPLICANT: Bjorn R. Olsen
TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/159,784
FILING DATE: December 1, 1993
CLASSIFICATION: 530
PRIOR APPLICATION NUMBER:
APPLICATION DATA:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: John F. Freeman
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00246/170001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8996
TELEX: 200154
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 195
TYPE: amino acid
STRANDEDNESS: N/A
TOPOLOGY: N/A
US-08-159-784-2

Query Match 100.0%; Score 41; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 0.61;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYLVICIE 8
DB 179 SYLVICIE 186

RESULT 3
US-09-206-059-2
Sequence 2, Application US/09206059
Patent No. 6201104
GENERAL INFORMATION:
APPLICANT: Macdonald, Nicholas
TITLE OF INVENTION: Angiogenesis-Inhibiting Protein Binding Peptides and
FILE REFERENCE: 05213-0370
CURRENT APPLICATION NUMBER: US/09/206,059
CURRENT FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 183
TYPE: PPT
ORGANISM: Homo sapiens
US-09-206-059-2

Query Match 92.7%; Score 38; DB 4; Length 183;
Best Local Similarity 87.5%; Pred. No. 2.1;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYLVICIE 8
DB 168 SYLVICIE 175

RESULT 4
US-09-191-647-10
Sequence 10, Application US/09191647
Patent No. 6046015
GENERAL INFORMATION:
APPLICANT: Goodman, Corey
APPLICANT: Kid, Thomas
APPLICANT: Brose, Katja
APPLICANT: Tessier-Lavigne, Marc
TITLE OF INVENTION: Modulating Robo: Ligand Interactions
FILE REFERENCE: 898-031-3
CURRENT APPLICATION NUMBER: US/09/191,647
CURRENT FILING DATE: 1998-11-13
EARLIER APPLICATION NUMBER: 60/065,544
EARLIER FILING DATE: 1997-11-14
EARLIER APPLICATION NUMBER: 60/081,057
EARLIER FILING DATE: 1998-04-07
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 10
LENGTH: 154
TYPE: PPT
ORGANISM: mouse
US-09-191-647-10

Query Match 73.2%; Score 30; DB 3; Length 154;
Best Local Similarity 62.5%; Pred. No. 60;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SYLVICIE 8
DB 44 SYLVICIE 51

RESULT 5
US-08-159-784-3
Sequence 3, Application US/08159784
Patent No. 5643783
GENERAL INFORMATION:
APPLICANT: Bjorn R. Olsen
TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/159,784
FILING DATE: December 1, 1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: John F. Freeman
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00246/170001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 191
TYPE: amino acid
STRANDEDNESS: N/A
TOPOLOGY: N/A
US-08-159-784-3

Query Match 73.2%; Score 30; DB 1; Length 191;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 IVLCIE 8
Db 177 IVLCIE 182

RESULT 6
US-08-948-616-9
Sequence 9, Application US/08948616
Patent No. 5840539
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purni
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
City: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/948,616
FILING DATE: Herevalth
CLASSIFICATION: 530
PRIOR APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0409 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 254 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 1223894

US-08-948-616-9

Query Match 73.2%; Score 30; DB 2; Length 254;
Best Local Similarity 62.5%; Pred. No. 99;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYVLCIE 8
Db 69 SYVLCIE 76

RESULT 7
US-09-193-510-9
Sequence 9, Application US/09193510
Patent No. 5981226
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purni
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
City: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/193,510
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/948,616
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0409 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 254 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 1223894
US-09-193-510-9

Query Match 73.2%; Score 30; DB 2; Length 254;
Best Local Similarity 62.5%; Pred. No. 99;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYVLCIE 8
Db 69 SYVLCIE 76

RESULT 8
US-09-368-402-9

Sequence 9, Application US/09368402
Patent No. 6200790
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purvi
TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/368,402
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/193,510
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0409 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 254 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1223894
US-09-368-402-9

Query Match 73.2%; Score 30; DB 4; Length 254;
Best Local Similarity 62.5%; Pred. No. 99;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYVLICIE 8
DB 69 SYMLICTE 76

RESULT 9
US-08-948-616-3
Sequence 3, Application US/08948616
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purvi
TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA

COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/948,616
FILING DATE: Herewith
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0409 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 307 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LUNGTUT07
CLONE: 2607662
US-08-948-616-3

Query Match 73.2%; Score 30; DB 2; Length 307;
Best Local Similarity 62.5%; Pred. No. 1.2e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYVLICIE 8
DB 69 SYMLICTE 76

RESULT 10
US-09-193-510-3
Sequence 3, Application US/09193510
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purvi
TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/193,510
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/948,616
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0409 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 307 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LUNGTUT07
CLONE: 2607662
US-09-193-510-3

Query Match
Best Local Similarity 73.2%; Score 30; DB 2; Length 307;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYVLCIE 8
11:1111
Db 69 SYMMLCTE 76

RESULT 11
US-09-368-402-3
Sequence 3, Application US/09368402
Patent No. 6200790
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Shah, Preeti
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/368,402
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/193,510
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0409 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 307 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

IMMEDIATE SOURCE:
LIBRARY: LUNGTUT07
CLONE: 2607662
US-09-368-402-3

Query Match
Best Local Similarity 73.2%; Score 30; DB 4; Length 307;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYVLCIE 8
11:1111
Db 69 SYMMLCTE 76

RESULT 12
US-08-723-938-3
Sequence 3, Application US/08723938
Patent No. 576759
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Coleman, Roger
TITLE OF INVENTION: TWO NOVEL HUMAN CATHESPIN PROTEINS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/723,938
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0125 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 395 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
IMMEDIATE SOURCE:
LIBRARY: LUNGTOT02
CLONE: 312099
US-08-723-938-3

Query Match
Best Local Similarity 73.2%; Score 30; DB 1; Length 395;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 YIVLCIE 8

Db 313 YIICSE 319

RESULT 13

US-09-080-538-3
Sequence 3, Application US/09080538
Patent No. 5965129
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Coleman, Roger
TITLE OF INVENTION: TWO NOVEL HUMAN CATHESPIN PROTEINS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/080,538
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/723,938
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0125 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 395 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
IMMEDIATE SOURCE:
LIBRARY: LUNGNOT02
CLONE: 312099
US-09-080-538-3

Query Match 73.2%; Score 30; DB 2; Length 395;
Best Local Similarity 71.4%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 YVICIE 8
DB 313 YIICSE 319

RESULT 14
US-09-008-271A-4
Sequence 4, Application US/09008271A
Patent No. 6203979
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
Hillman, Jennifer L.

Yue, Henry
Guegler, Karl J.
Corley, Neil C.
Tang, Tom Y.
Shah, Purvi
TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/008,271A
FILING DATE: 16-Jan-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mohan-Peterson, Sheela
REGISTRATION NUMBER: 41,201
REFERENCE/DOCKET NUMBER: PF-0458 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 420 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LUNGAST01
CLONE: 877617
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-008-271A-4

Query Match 73.2%; Score 30; DB 4; Length 420;
Best Local Similarity 71.4%; Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 YVICIE 8
DB 313 YIICSE 319

RESULT 15
US-08-974-691-8
Sequence 8, Application US/08974691
Patent No. 6225103
GENERAL INFORMATION:
APPLICANT: Keolsch, Gerald
APPLICANT: Lin, Xinli
APPLICANT: Tang, Jordan
TITLE OF INVENTION: Cloning and Characterization of Napsin
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center, 1201 W. Peachtree
STREET: St.
CITY: Atlanta
STATE: GA
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,691
FILING DATE: 20-NOV-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/031,196
FILING DATE: 20-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/046,126
FILING DATE: 09-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patricia L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: OMR# 166
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-873-8794
TELEFAX: 404-873-8795
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 420 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-974-691-8

Query Match 73.2%; Score 30; DB 4; Length 420;
Best Local Similarity 71.4%; Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 YIVLCIE 8
||:|:|:|
Db 313 YIILCSE 319

Search completed: August 16, 2001, 11:41:24
Job time: 407 sec

PT Mutant endostatin having anti-angiogenic activity

XX Claim 1; Page 71; 105pp; English.

XX This sequence represents the C-terminalus from the mutant endostatin (EM)
 CC of the invention, which has anti-angiogenic activity, and is designated
 CC EMI. Compositions comprising EMI or fusion proteins comprising EMI, are
 CC useful for treating diseases characterised by angiogenic activity, such
 CC as angiogenesis-dependent cancers, benign tumours, rheumatoid arthritis,
 CC psoriasis, ocular angiogenesis, Osler-Webber Syndrome, myocardial
 CC angiogenesis, plaque neovascularisation, telangiectasia, haemophilic
 CC joints, angiodiroma, wound granulation, intestinal adhesions,
 CC atherosclerosis, scleroderma, hypertrophic scars, cat scratch disease,
 CC Helicobacter pylori ulcers, dialysis graft vascular access stenosis,
 CC contraception and obesity. In particular, the diseases treatable by EMI
 CC comprising cancer, especially renal cancer. The methods provide a means for
 CC introducing EMI into mammalian cells via gene therapy, for production of
 CC EMI via recombinant means, as well as recombinant production of the EMI
 CC protein. EMI performs as well or better than whole endostatin. Use
 CC of EMI is advantageous for treatment of angiogenic diseases in that
 CC increasingly smaller peptides are more potent on a weight basis, and may
 CC be able to better penetrate tissues.

CC Sequence 8 AA:

Query Match 100.0%; Score 41; DB 20; Length 8;
 Best Local Similarity 100.0%; Pred. No. 3 4e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYLVICIE 8

Db 1 sylvicic 8

RESULT 2

AAB49808
 ID AAB49808 standard; Protein; 48 AA.

AC AAB49808;

DT 02-MAR-2001 (first entry)

DE Murine endostatin peptide fragment SEQ ID NO: 21.

KW Endostatin; antiangiogenic; angiogenesis; human; mouse; chicken;
 KM cancer; inflammation; angiogenesis-dependent disease.

XX Mus musculus.

OS WO200067771-A1.

PN 16-NOV-2000.

PD 02-MAY-2000; 2000WO-US12063.

PF 06-MAY-1999; 99US-0132907.

PR 14-JUL-1999; 99US-0353333.

XX (BURN-) BURNHAM INST.

XX VUori K;

XX WPI; 2001-040937/05.

XX Endostatin peptide comprising at least four endostatin amino acid
 PT residues are e.g. angiogenesis inhibitors for treating cancer and
 PT diabetic retinopathy

PS Claim 10; Page 124-125; 146pp; English.

CC The present invention provides endostatin peptides which can be used in
 CC the modulation of angiogenesis. This is useful in the treatment of

CC cancers, inflammation, rheumatoid arthritis, chronic articular
 CC rheumatism, psoriasis, disorders associated with inopportune invasion of
 CC vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy
 CC of prematurity, macular degeneration, corneal graft rejection,
 CC retrolental fibroplasia, rubeosis, capillary proliferation in
 CC atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent
 CC diseases include Osler-Webber syndrome, myocardial angiogenesis, plaque
 CC neovascularisation, telangiectasia, haemophilic joints and wound
 CC granulation. In addition, the peptides can be used as birth control
 CC agents.

CC Sequence 48 AA:

Query Match 100.0%; Score 41; DB 22; Length 48;
 Best Local Similarity 100.0%; Pred. No. 0.25;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYLVICIE 8

Db 37 sylvicic 44

RESULT 3

AAB35587
 ID AAB35587 standard; Peptide; 50 AA.

AC AAB35587;

DT 14-FEB-2001 (first entry)

DE Antiangiogenic pentacontapeptide IV.

KW Antiangiogenic; angiogenesis; cancer; endostatin.

OS Synthetic.

PN WO200063249-A1.

PD 26-OCT-2000.

PF 11-APR-2000; 2000WO-EP03236.

PR 15-APR-1999; 99IT-MI00777.

XX (UWMI-) UNITV MILANO.

PA (UWFI-) UNITV FIRENZE.

XX Chillemi F, Francescato P, Ziche M;

DR WPI; 2001-007005/01.

XX Polypeptides derived from endostatin exhibiting antiangiogenic activity
 PT useful for treatment of angiogenesis-dependent tumours

PS Claim 5; Page 18; 28pp; English.

CC The present invention describes a number of peptides derived from
 CC endostatin which exhibit antiangiogenic activity. These may be used in
 CC the treatment of cancer. The present sequence is one of the peptides of
 CC the invention.

CC Sequence 50 AA:

Query Match 100.0%; Score 41; DB 22; Length 50;
 Best Local Similarity 100.0%; Pred. No. 0.26;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYLVICIE 8

Db 34 sylvicic 41

RESULT 4
AAV18409
ID AAV18409 standard; Protein: 184 AA.
XX
AC AAV18409;
XX
DT 24-AUG-1999 (first entry)
XX
DE Endostatin protein sequence.
XX
KW EMI; anti-angiogenic peptide; endostatin; angiogenesis-dependent cancer;
KW benign tumour; rheumatoid arthritis; psoriasis; ocular angiogenesis;
KW Osler-Webber Syndrome; myocardial angiogenesis; angiolipoma; cancer;
KW plaque neovascularisation; telangiectasia; atherosclerosis; scleroderma;
KW dialysis graft vascular access stenosis; renal cancer; therapy.
XX
OS Mus sp.
XX
PN WO929855-A1.
XX
PD 17-JUN-1999.
XX
PF 08-DEC-1998; 98MO-US26057.
XX
PR 16-NOV-1998; 98US-0108536.
PR 08-DEC-1997; 97US-0067888.
PR 22-APR-1998; 98US-0082663.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PT Sukhatme VP;
XX
DR MPI; 1999-385604/32.
XX
DR N-PSDB; AAX79949.
XX
PT Mutant endostatin having anti-angiogenic activity
XX
PS Claim 31; Fig 2; 105pp; English.
XX
CC This sequence is the mouse endostatin. The invention relates to a
CC the mutant endostatin (EMI), which has anti-angiogenic activity, and is
CC designated EMI. Compositions comprising EMI or fusion proteins comprising
CC EMI, are useful for treating diseases characterised by angiogenic
CC activity, such as angiogenesis-dependent cancers, benign tumours,
CC rheumatoid arthritis, psoriasis, ocular angiogenesis, Osler-Webber
CC Syndrome, myocardial angiogenesis, plaque neovascularisation,
CC telangiectasia, haemophilic joints, angiolipoma, wound granulation,
CC intestinal adhesions, atherosclerosis, scleroderma, hypertrophic scars,
CC cat scratch disease, Helicobacter pylori ulcers, dialysis graft vascular
CC access stenosis, contraception and obesity. In particular, the diseases
CC treatable by EMI comprise cancer, especially renal cancer. The methods
CC provide a means for introducing EMI into mammalian cells via gene
CC therapy, for production of EMI via recombinant means, as well as
CC recombinant production of the EMI protein. EMI performs as well or better
CC than whole endostatin. Use of EMI is advantageous for treatment of
CC angiogenic diseases in that increasingly smaller peptides are more potent
CC on a weight basis, and may be able to better penetrate tissues.
XX
SO Sequence 184 AA;

Query Match 100.0%; Score 41; DB 20; Length 184;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCIE 8
 |||||
Db 168 syivlcie 175

RESULT 5
AAV08689
ID AAV08689 standard; Protein: 184 AA.
XX
AC AAV08689;
XX
DT 10-AUG-1999 (first entry)
XX
DE Murine endostatin protein fragment.
XX
KW Plasminogen; murine; angiotatin; endostatin; gene therapy; vector;
KW anti-angiogenic; attenuator; cytosolic; anti-diabetic; ophthalmology;
KW tumour growth; solid tumour; diabetic retinopathy; retina.
XX
OS Mus sp.
XX
PN WO926480-A1.
XX
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98MO-US24950.
XX
PR 20-NOV-1997; 97US-0975424.
XX
PA (GENE-) GENETIX PHARM INC.
PA (MAST) MASSACHUSETTS INST TECHNOLOGY.
XX
PT Bachelot T, Leboulch P, Pawluc RJ;
XX
DR MPI; 1999-357696/30.
XX
DR N-PSDB; AAX77715.
XX
PT Anti-angiogenic gene therapy vectors
XX
PS Disclosure; Fig 6; 83pp; English.
XX
CC This invention describes a novel viral gene therapy vector comprising a
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen
CC from human or murine angiotatin, human or murine endostatin and
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is
CC sufficiently attenuated for use in human gene therapy. The products of
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and
CC ophthalmological activity. The vector is used in gene therapy for
CC inhibiting tumour growth in humans harbouring a solid tumour. The vector
CC expresses an anti-angiogenic polypeptide. An additional use comprises
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide
CC inhibits angiogenesis in the vicinity of the retina. The vector is
CC administered to cells ex vivo and then administered to the patient.
XX
SO Sequence 184 AA;

Query Match 100.0%; Score 41; DB 20; Length 184;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCIE 8
 |||||
Db 168 syivlcie 175

RESULT 6
AAV70258
ID AAV70258 standard; Protein: 184 AA.
XX
AC AAV70258;
XX
DT 06-JUN-2000 (first entry)
XX
DE Murine angiogenesis inhibitor, endostatin.
XX
KW Murine; immunoglobulin Fc fragment; endostatin; immunofusin;
KW angiogenesis; inhibitor; cytosolic; antirheumatoid; antiarthritic;
KW antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;
KW vasotropic; vulnary; treatment; antiarteriosclerosis; tumour;
XX

KM metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;
 KM ocular angiogenic disease; diabetic retinopathy; macular degeneration;
 KM myocardial angiogenesis; plaque neovascularisation; telangiectasia;
 KM wound granulation; keloid scar; gene therapy.
 OS Mus musculus.
 PN W0200011033-A2.
 XX
 XX
 PD 02-MAR-2000.
 XX
 XX 25-AUG-1999; 99MO-US19329.
 PE 25-AUG-1998; 98US-0097883.
 PR 25-AUG-1998; 98US-0097883.
 XX
 XX (LEXI-) LEXINGEN PHARM CORP.
 PA
 XX
 XX Lo K, Li Y, Gillies SD;
 PI
 DR MPI; 2000-237616/20.
 DR N-PSDB; AAZ51299.
 XX
 PT Novel fusion protein of angiotensin or endostatin and an immunoglobulin
 PT FC region, useful for treating conditions mediated by angiogenesis,
 PT such as rheumatoid arthritis, tumors and macular degeneration -
 PS
 XX Example 5; Pages 48-49; 68pp; English.
 CC The patent discloses a DNA molecule encoding a fusion protein comprising
 CC a signal sequence, an immunoglobulin FC region, and an angiogenesis
 CC inhibitor selected from angiotensin, endostatin, a plasminogen fragment
 CC having angiotensin activity, a collagen XVIII fragment having endostatin
 CC activity, or combinations of them. The fusion protein (immunofusin) is
 CC used to inhibit angiogenesis and to treat diseases or conditions mediated
 CC by angiogenesis. Conditions that may be treated include solid tumours,
 CC blood born tumours, tumour metastasis, benign tumours including
 CC haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic
 CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases
 CC e.g. diabetic retinopathy, retinopathy of prematurity, macular
 CC degeneration, corneal graft rejection, neovascular glaucoma, retrolental
 CC fibroplasia, rubeosis and Osler-Weber syndrome, myocardial angiogenesis,
 CC plaque neovascularisation, telangiectasia, haemophilic joints,
 CC angiodioma, wound granulation, and excessive or abnormal stimulation of
 CC endothelial cells, intestinal cells, atherosclerosis, sclerodermal and
 CC hypertrophic scars, i.e. keloid scars. The DNA constructs may be used
 CC in gene therapy. The present sequence is a murine
 CC endostatin used in the construction of immunofusin containing murine
 CC immunoglobulin FC fragment.
 CC
 XX
 SQ Sequence 184 AA:
 Query Match 100.0%; Score 41; DB 21; Length 184;
 Best Local Similarity 100.0%; Pred. No. 0.89;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 STIVLCIE 8
 IIIIIIII
 DB 168 sylvic1e 175
 RESULT 7
 AAB49380
 ID AAB49380 standard; Protein: 184 AA.
 XX
 XX AAB49380;
 AC
 XX
 XX 02-MAR-2001 (first entry)
 DT
 XX
 XX Murine endostatin SEQ ID NO: 4.
 DE
 XX
 XX Endostatin: antiangiogenic; angiogenesis; human; mouse; chicken;
 KM cancer; inflammation; angiogenesis-dependent disease.

XX
 OS Mus musculus.
 XX
 XX
 PN W0200067771-A1.
 XX
 XX
 PD 16-NOV-2000.
 XX
 XX
 PE 02-MAY-2000; 2000MO-US12063.
 XX
 XX 06-MAY-1999; 99US-0132907.
 PR 14-JUL-1999; 99US-0353333.
 XX
 XX (BURN-) BURNHAM INST.
 PA
 XX
 XX Vuori K;
 PI
 DR MPI; 2001-040937/05.
 DR N-PSDB; AAC88290.
 XX
 PT Endostatin peptide comprising at least four endostatin amino acid
 PT residues are e.g. angiogenesis inhibitors for treating cancer and
 PT diabetic retinopathy -
 PS
 XX Disclosure; Fig 1; 146pp; English.
 CC The present invention provides endostatin peptides which can be used in
 CC the modulation of angiogenesis. This is useful in the treatment of
 CC cancers, inflammation, rheumatoid arthritis, chronic articular
 CC rheumatism, psoriasis, disorders associated with inopportune invasion of
 CC vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy
 CC of prematurity, macular degeneration, corneal graft rejection,
 CC retrolental fibroplasia, rubeosis, capillary proliferation in
 CC atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent
 CC diseases include Osler-Weber syndrome, myocardial angiogenesis, plaque
 CC neovascularisation, telangiectasia, haemophilic joints and wound
 CC granulation. In addition, the peptides can be used as birth control
 CC agents.
 CC
 XX
 SQ Sequence 184 AA:
 Query Match 100.0%; Score 41; DB 22; Length 184;
 Best Local Similarity 100.0%; Pred. No. 0.89;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 STIVLCIE 8
 IIIIIIII
 DB 168 sylvic1e 175
 RESULT 8
 AAY06197
 ID AAY06197 standard; Protein: 185 AA.
 XX
 XX AAY06197;
 AC
 XX
 XX 16-AUG-1999 (first entry)
 DT
 XX
 XX Anti-angiogenic endostatin peptide.
 DE
 XX
 XX Anti-angiogenic; carrier:DNA complex; tumour; gene therapy; human;
 KM endostatin; melanoma; lung cancer; colon cancer; brain cancer;
 KM breast cancer.
 XX
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 36
 FT /note- "encoded by CAG"
 FT Misc-difference 37
 FT /note- "encoded by CAA"
 FT Misc-difference 39
 FT /note- "deduced sequence from nucleotide sequence

	FT	Misc-difference	76	/note-	"encoded by AAG"	
	FT	Misc-difference	118	/note-	"encoded by AAG"	
	FT	Misc-difference	162	/note-	"encoded by AAA"	
	FT	Misc-difference	168	/note-	"encoded by AAC"	
	FT	Misc-difference	185	/note-	"encoded by AAA"	
	PB	EPR921193-AI.				
	PD	09-JUN-1999.				
	PE	07-JAN-1998;	98EP-0100135.			
	PF	05-DEC-1997;	97US-0985526.			
	PG	(MIXS/) MIXSON A J.				
	PH	Mixson AJ:				
	PI	WPI; 1999-315406/27.				
	PL	N-PADB; AAAS58740.				
	PM	Inhibition of growth of solid tumors				
	PN	Disclosure; Page 38; 46pp; English.				
	PO	The present sequence represents an anti-angiogenic endostatin peptide. The invention provides a carrier-DNA complex that comprises DNA (see AAAS8725-42) encoding an anti-angiogenic protein or polypeptide, such as the present sequence, the complex being deliverable to the site of a tumour in vivo, and additionally comprising regulatory elements for expressing the anti-angiogenic DNA in a tumour or tumour vasculature. The complex may also include DNA encoding a liposome suppressor protein, especially p53. The carrier is a component, or a combination of these, and administration is by intravenous or intra-tumoral injection. The complexes are useful in gene therapy for inhibition of tumour growth. The types of tumors which may be treated include solid tumors such as melanomas and tumors in the lung, colon, brain and breast.				
	PP	Sequence	185 AA;			
	PQ	Query Match	100.0%; Score 41; DB 20; Length 185;			
	PR	Best Local Similarity	100.0%; Pred. No. 0.89;			
	PS	Matches	8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
	PT	1 SYIVLCIE	8			
	PV	DB	169 syivlcie	176		
	PX	RESULT	9			
	PY	AAB28398	standard; Protein; 191 AA.			
	PZ	AAAC	AAB28398;			
	QA	DT	19-FEB-2001 (first entry)			
	QB	DE	Murine endostatin.			
	QC	KW	Murine endostatin; cytotoxic; antiproliferative;			
	QE	KW	vascular endothelial growth factor; VEGF; antibody; VEGFR2 receptor;			
	QF	KW	cancer; vascularised solid tumor.			
	QG	XX				

```

XX Mus sp.
XX PN WO200064946-A2.
XX PD 02-NOV-2000.
XX PF 28-APR-2000; 2000WO-US11367.
XX PR 28-APR-1999; 99US-0131432.
XX PA (TEXA ) UNIV TEXAS SYSTEM.
XX PI Thorpe PE, Brekken RA;
XX N-PSDB: AAC67777.
XX WPI: 2000-687317/67.
XX N-PSDB: AAC67777.
XX PF Immunogenic composition for the treatment and diagnosis of cancer
XX PT comprises an anti-VEGF (vascular endothelial growth factor) antibody
XX PT binding the same epitope as the monoclonal antibody ATCC PTA 1595 -
XX PS Example 10; Page 290-291; 298pp; English.
XX CC The present invention relates to anti-Vascular Endothelial Growth Factor
XX CC (VEGF) antibodies that bind to the same epitope as the monoclonal
XX CC antibody ATCC PTA 1595 and which significantly inhibit VEGF binding to
XX CC the VEGF receptor VEGFR2, without inhibiting VEGF binding to the VEGF
XX CC receptor VEGFR1. The present sequence is murine endostatin. Endostatin
XX CC may be conjugated onto the anti-VEGF antibodies of the present invention.
XX CC The anti-VEGF antibodies of the present invention are useful for the
XX CC treatment and diagnosis of cancer, especially vascularised solid tumours.
XX SQ Sequence 191 AA:

Query Match 100.0%; Score 41; DB 21; Length 191;
Best Local Similarity 100.0%; Pred. No. 0.92;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

Qy 1 SYLVICIE 8
|||
Db 175 SYLVICIE 182

RESULT 10
AAB71930
ID AAB71930 standard; Protein; 207 AA.
XX AC AAB71930;
XX DT 10-MAY-2001 (first entry)
XX DE Murine endostatin attached to Ig-kappa leader sequence.
XX KW Mouse; endostatin; antitumour; cytostatic; antiarthritic; antipsoriatic;
XX KW antidiabetic; ophthalmological; gene therapy; angiogenic inhibitor;
XX KW adenoviral vector; diabetic retinopathy; cardiovascular disease;
XX KW arthritis; psoriasis; cerebral oedema; intravascular coagulopathy;
XX KW lymphoma; leukaemia; immunoglobulin; Ig; Ig-kappa.
XX OS Mus sp.
XX PN WO200112830-A1.
XX PD 22-FEB-2001.
XX PF 11-AUG-2000; 2000WO-EP07865.
XX PR 13-AUG-1999; 99US-0373938.
XX PA (NOVS ) NOVARTIS AG.
XX PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX

```

PI Hallenbeck PL, Chen CT;
XX
XX WPI: 2001-202871/20.
DR N-PSDB; AAF60336.
XX
XX Adenoviral vector for treating tumors and disorders associated with
PT angiogenesis, such as cancer, arthritis, and psoriasis, comprises a DNA
PT sequence encoding an angiogenic inhibitor, particularly endostatin
XX
XX Example 1; Fig 1B; 59pp; English.
PS
CC The nucleotide sequence encoding this protein was used in the
CC construction of an adenoviral vector which includes a DNA sequence
CC encoding endostatin. The adenoviral vector is useful for expressing
CC endostatin in a mammalian cell such as an A549 or Hep3b cell. It is
CC useful for treating other diseases and disorders associated with
CC angiogenesis, such as neovascular diseases and disorders associated with
CC retinopathy, cardiovascular disease, arthritis, psoriasis, cerebral
CC edema and intravascular coagulopathy (Kasabach-Merritt syndrome). The
CC vector inhibits, prevents or destroys the growth of tumours by
CC preventing the formation of blood vessels in tumours, such as lymphoma
CC and leukaemia.
CC
SO Sequence 207 AA:

Query Match 100.0%; Score 41; DB 22; Length 207;
Best Local Similarity 100.0%; Pred. No. 0.99;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYVLICIE 8
DB 191 syvlicie 198
|||||||
RESULT 11
AA08691
ID AAY08691 standard; Protein; 218 AA.
XX
XX AAY08691;
AC
XX
DT 10-AUG-1999 (first entry)
XX
DE Murine gene therapy peptide construct SP-Flag-Endo.
XX
XX Plasmidogen; murine; angiotatin; endostatin; gene therapy; vector;
KW anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
KW tumour growth; solid tumour; diabetic retinopathy; retina; construct.
XX
XX Mus sp.
OS Synthetic.
OS
XX W09926480-A1.
XX
XX 03-JUN-1999.
PD
XX 20-NOV-1998; 98WO-US24950.
PF
XX 20-NOV-1997; 97US-0975424.
PR
XX
PA (GENE-) GENETIX PHARM INC.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
XX
XX Bachelot T, Leboulch P, Pawllok RJ;
XX
XX WPI: 1999-357696/30.
DR N-PSDB; AAX77717.
XX
XX Anti-angiogenic gene therapy vectors
XX
XX Example 1; Page 69; 83pp; English.
XX
CC This invention describes a novel viral gene therapy vector comprising a

CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen
CC from human or murine angiotatin, human or murine endostatin and
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is
CC sufficiently attenuated for use in human gene therapy. The products of
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and
CC ophthalmological activity. The vector is used in gene therapy for
CC inhibiting tumour growth in humans harbouring a solid tumour. The vector
CC expresses an anti-angiogenic polypeptide. An additional use comprises
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide
CC inhibits angiogenesis in the vicinity of the retina. The vector is
CC administered to cells ex vivo and then administered to the patient.
XX
SO Sequence 218 AA:

Query Match 100.0%; Score 41; DB 20; Length 218;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYVLICIE 8
DB 202 syvlicie 209
|||||||
RESULT 12
AA08692
ID AAY08692 standard; Protein; 580 AA.
XX
XX AAY08692;
AC
XX
DT 10-AUG-1999 (first entry)
XX
DE Murine gene therapy peptide construct SP-K1-K2-K3-K4-Flag-Endo.
XX
XX Plasmidogen; murine; angiotatin; endostatin; gene therapy; vector;
KW anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
KW tumour growth; solid tumour; diabetic retinopathy; retina; construct.
XX
XX Mus sp.
OS Synthetic.
OS
XX W09926480-A1.
XX
XX 03-JUN-1999.
PD
XX 20-NOV-1998; 98WO-US24950.
PF
XX 20-NOV-1997; 97US-0975424.
PR
XX
PA (GENE-) GENETIX PHARM INC.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
XX
XX Bachelot T, Leboulch P, Pawllok RJ;
XX
XX WPI: 1999-357696/30.
DR N-PSDB; AAX77718.
XX
XX Anti-angiogenic gene therapy vectors
XX
XX Example 1; Page 72-74; 83pp; English.
XX
CC This invention describes a novel viral gene therapy vector comprising a
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen
CC from human or murine angiotatin, human or murine endostatin and
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is
CC sufficiently attenuated for use in human gene therapy. The products of
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and
CC ophthalmological activity. The vector is used in gene therapy for
CC inhibiting tumour growth in humans harbouring a solid tumour. The vector
CC expresses an anti-angiogenic polypeptide. An additional use comprises
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide
CC inhibits angiogenesis in the vicinity of the retina. The vector is
CC administered to cells ex vivo and then administered to the patient.

XX
SQ Sequence 580 AA;

Query Match 100.0%; Score 41; DB 20; Length 580;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYVLICIE 8
Db 564 SYVLICIE 571

RESULT 13

AAV25114
ID AAV25114 standard; Protein; 684 AA.

XX
AC AAV25114;

XX
DT 25-AUG-1999 (first entry)

XX
DE Mouse alpha1 (XVIII) collagen protein.

XX
KW Alpha1(XVIII) collagen; mimetic; endostatin; atomic coordinate; library;
KW anti-angiogenic; heparin binding domain; receptor binding domain; mimic;
KW alpha-helix A domain; carbohydrate recognition domain; CRD domain;
KW treatment; angiogenesis; tumour; murine.

XX
OS Mus sp.

XX
PN WO9931616-A1.

XX
PD 24-JUN-1999.

XX
PF 16-DEC-1998; 98MO-US26783.

XX
PR 16-DEC-1997; 97US-0069727.

XX
PA (HARD) HARVARD COLLEGE.

XX
PI Hohenester E, Olsen BR, Sasaki T, Timpl R;

XX
DR WPI; 1999-395243/33.

XX
PT Identifying mimetics of mammalian endostatin

XX
PS Disclosure; Fig 5A-C; 75pp; English.

XX
CC This invention describes a novel method for identifying mimetics of
CC mammalian endostatin. The method comprises identifying a compound
CC having atomic coordinates with non-trivial similarity to selected
CC coordinates of atoms of a mammalian endostatin involves (a) providing
CC a library of atomic coordinates of compounds in a library of candidate
CC compounds, (b) comparing the library of atomic coordinates to the
CC selected coordinates of a mammalian endostatin and (c) selecting from the
CC library at least one candidate compound on the basis of selection
CC criteria which include similarities between the atomic coordinates of the
CC selected candidate compound and the atomic coordinates of the mammalian
CC endostatin. The invention also describes the use of an anti-angiogenic
CC fragment of endostatin comprising a domain selected from a heparin
CC binding domain, a receptor binding domain, and exposed on alpha-helix A
CC domain, and a carbohydrate recognition domain (CRD) domain. The methods
CC can be used for designing and selecting endostatin mimics. The compounds
CC identified can be used for treating undesired angiogenesis, e.g. tumours.
CC This sequence represents mouse alpha1(XVIII) collagen which is used in
CC the description of the method.

XX
SQ Sequence 684 AA;

Query Match 100.0%; Score 41; DB 20; Length 684;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYVLICIE 8
Db 669 SYVLICIE 676

RESULT 14

AAW26328
ID AAW26328 standard; Protein; 1288 AA.

XX
AC AAW26328;

XX
DT 19-NOV-1997 (first entry)

XX
DE Mouse alpha-1 collagen (XVIII).

XX
KW Alpha-1 collagen; type XVIII collagen; cartilage degeneration.

XX
OS Mus musculus.

XX
PN Key

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

XX
FT Peptide

```

FT Peptide 506..511 /label= GXYGX'Y'_motif
FT Peptide 512..517 /label= GXYGX'Y'_motif
FT Peptide 518..523 /label= GXYGX'Y'_motif
FT Peptide 524..529 /label= GXYGX'Y'_motif
FT Peptide 530..535 /label= GXYGX'Y'_motif
FT Peptide 536..541 /label= GXYGX'Y'_motif
FT Peptide 542..547 /label= GXYGX'Y'_motif
FT Peptide 548..553 /label= GXYGX'Y'_motif
FT Peptide 580..585 /label= GXYGX'Y'_motif
FT Peptide 586..591 /label= GXYGX'Y'_motif
FT Peptide 592..597 /label= GXYGX'Y'_motif
FT Peptide 598..603 /label= GXYGX'Y'_motif
FT Peptide 604..609 /label= GXYGX'Y'_motif
FT Peptide 610..615 /label= GXYGX'Y'_motif
FT Peptide 616..621 /label= GXYGX'Y'_motif
FT Peptide 622..627 /label= GXYGX'Y'_motif
FT Peptide 628..633 /label= GXYGX'Y'_motif
FT Peptide 634..639 /label= GXYGX'Y'_motif
FT Peptide 640..665 /label= GXYGX'Y'_motif
FT Peptide 657..662 /label= GXYGX'Y'_motif
FT Peptide 677..682 /label= GXYGX'Y'_motif
FT Peptide 683..688 /label= GXYGX'Y'_motif
FT Peptide 689..694 /label= GXYGX'Y'_motif
FT Peptide 695..700 /label= GXYGX'Y'_motif
FT Peptide 707..712 /label= GXYGX'Y'_motif
FT Peptide 713..718 /label= GXYGX'Y'_motif
FT Peptide 735..740 /label= GXYGX'Y'_motif
FT Peptide 741..746 /label= GXYGX'Y'_motif
FT Peptide 747..752 /label= GXYGX'Y'_motif
FT Peptide 759..764 /label= GXYGX'Y'_motif
FT Peptide 765..770 /label= GXYGX'Y'_motif
FT Peptide 771..776 /label= GXYGX'Y'_motif
FT Peptide 787..792 /label= GXYGX'Y'_motif
FT Peptide 793..798 /label= GXYGX'Y'_motif
FT Peptide 799..804 /label= GXYGX'Y'_motif
FT Peptide 815..820 /label= GXYGX'Y'_motif
FT Peptide 821..826 /label= GXYGX'Y'_motif

```

```

FT Peptide /label= GXYGX'Y'_motif
FT Peptide 827..832 /label= GXYGX'Y'_motif
FT Peptide 833..838 /label= GXYGX'Y'_motif
FT Peptide 839..844 /label= GXYGX'Y'_motif
FT Peptide 845..850 /label= GXYGX'Y'_motif
FT Peptide 853..868 /label= GXYGX'Y'_motif
FT Peptide 869..874 /label= GXYGX'Y'_motif
FT Peptide 875..880 /label= GXYGX'Y'_motif
FT Peptide 891..896 /label= GXYGX'Y'_motif
FT Peptide 897..902 /label= GXYGX'Y'_motif
FT Peptide 903..908 /label= GXYGX'Y'_motif
FT Peptide 911..916 /label= GXYGX'Y'_motif
FT Peptide 917..922 /label= GXYGX'Y'_motif
FT Peptide 928..933 /label= GXYGX'Y'_motif
FT Peptide 934..939 /label= GXYGX'Y'_motif
FT Peptide 956..961 /label= GXYGX'Y'_motif
FT Peptide 962..967 /label= GXYGX'Y'_motif
FT Peptide 968..973 /label= GXYGX'Y'_motif
FT Peptide 1126..1131 /label= GXYGX'Y'_motif
FT Peptide 1145..1150 /label= GXYGX'Y'_motif
FT Peptide 1193..1198 /label= GXYGX'Y'_motif
FT Peptide /label= GXYGX'Y'_motif

```

```

XX US5643783-A.
XX
XX
XX PD 01-JUL-1997.
XX
XX PF 01-DEC-1993; 93US-0159784.
XX
XX PR 01-DEC-1993; 93US-0159784.
XX
XX PA (HARD ) HARVARD COLLEGE.
XX
XX PI Oh SP, Olsen BR:
XX
XX DR MPI: 1997-350247/32.
XX DR N-PSDB: AAT84485.
XX
XX Nucleic acid encoding human alpha-1 collagen - for production of
PT recombinant alpha-1 collagen, for use in the treatment of cartilage
PT degeneration
XX
XX PS Disclosure; Fig 2; 35pp; English.

```

Query Match 100.0%; Score 41; DB 18; Length 1288;
 Best Local Similarity 100.0%; Pred. No. 5.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SYIVLCIE 8
 |||||
 DB 1272 syivlcie 1279

RESULT 15


```

AAW92297
ID   AAW92297 standard; peptide; 1288 AA.
XX
AC   AAW92297;
XX
DT   28-APR-1999 (first entry)
XX
DE   Mouse alpha-1 (XVIII) collagen chain common sequence MO18(common)28.
XX
KW   Human; type XVIII collagen; liver disease; cirrhosis; detection;
XX   hepatocellular carcinoma; diagnosis.
XX
OS   Mus sp.
XX
PN   MO9856399-A1.
XX
PD   17-DEC-1998.
XX
PE   12-JUN-1998; 98MO-US12327.
XX
PR   12-JUN-1997; 97US-0049369.
XX
PA   (FIFT-) ACAD FINLAND.
PA   (FIBR-) FIBROGEN INC.
PA   (INRM ) INST NAT SANTE & RECH MEDICALE.
XX
PI   Clement B, Pihlajaniemi T, Rehn M;
XX
DR   WPI; 1999-070292/06.
XX
PT   Diagnosis and monitoring of liver disease by measuring collagen type
PT   XVIII levels - with elevated levels indicative of disease,
PT   especially cirrhosis or hepatocellular carcinoma
XX
PS   Example 6; Fig 8; 56pp; English.
XX
CC   A method has been developed for the detecting liver disease. The method
CC   comprises: (a) reacting a patient sample with antibodies (Ab) specific
CC   for collagen type XVIII (Coll18); (b) measuring the amount of Ab-antigen
CC   complex (C) formed as indicator of the amount of Coll18 present; (c)
CC   similar analysis of a non-diseased control; and (d) comparing the
CC   amounts of Coll18 in the two samples to detect presence or progression of
CC   disease. Elevated levels of Coll18 are: (i) indicative of disease,
CC   specifically cirrhosis; and (ii) predictive of the prognosis of disease,
CC   specifically hepatocellular carcinoma (there is a relationship between
CC   Coll18 mRNA levels and tumour size and necrosis, and survival times are
CC   significantly higher in patients with higher Coll18 levels). The method
CC   provides non-invasive, early and accurate diagnosis of liver disease.
CC   The present sequence represents the sequence common to mouse alpha-1
CC   (XVIII) collagen chain from the present invention.
XX
SQ   Sequence 1288 AA:

```

```

Query Match      100.0%; Score 41; DB 20; Length 1288;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY   1 SYIVICIE 8
Db   1273 syivicle 1280

```

Search completed: August 16, 2001, 11:41:06
Job time: 504 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 11:41:07 ; Search time 21.98 Seconds
(without alignments)
48.155 Million cell updates/sec

Title: US-09-589-777A-2_COPY_168_175

Perfect score: 41

Sequence: 1 SYIVLCIE 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: SP:REMBL.16:*
2: SP:archaea:*
3: SP:bacteria:*
4: SP:fungi:*
5: SP:human:*
6: SP:invertebrate:*
7: SP:mammal:*
8: SP:mhc:*
9: SP:organelle:*
10: SP:phage:*
11: SP:plant:*
12: SP:rodent:*
13: SP:unclassified:*
14: SP:vertebrate:*
15: SP:virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	171	11	09WUW5
2	41	100.0	184	11	09K63
3	41	100.0	226	11	0902D2
4	41	100.0	1140	11	061434
5	41	100.0	1774	11	062001
6	33	80.5	237	2	087543
7	33	80.5	386	13	093419
8	33	80.5	646	5	015900
9	33	80.5	706	5	015901
10	33	80.5	1905	5	09XTP6
11	32	78.0	147	2	084569
12	32	78.0	147	2	09P3H5
13	32	78.0	347	1	09P9G6
14	32	78.0	507	3	09P905
15	32	78.0	613	4	09S942
16	32	78.0	1077	13	091019
17	32	78.0	1142	13	002391
18	32	78.0	1160	11	0921E9
19	32	78.0	1171	11	062638

20	32	78.0	1175	11	061543	061543 mus musculus
21	32	78.0	1177	4	013221	013221 homo sapien
22	32	78.0	1179	4	092896	092896 homo sapien
23	31	75.6	70	4	09P1N2	09P1N2 homo sapien
24	31	75.6	158	3	092784	092784 chlamydia p
25	31	75.6	163	10	09SLC0	09SLC0 arbidolopis
26	31	75.6	213	14	09J5B0	09J5B0 fowlpox vir
27	31	75.6	235	5	093866	093866 caenorhabd1
28	31	75.6	272	14	09BN36	09BN36 ansacta moo
29	31	75.6	316	2	09RST2	09RST2 dehnococcus
30	31	75.6	466	2	09KMS0	09KMS0 vibrio chol
31	31	75.6	472	5	09VON6	09VON6 drosophila
32	31	75.6	479	10	09X152	09X152 arbidolopis
33	31	75.6	566	2	09PHU2	09PHU2 campylobact
34	31	75.6	991	5	095024	095024 dictyostell
35	31	75.6	2158	10	09LUT5	09LUT5 arbidolopis
36	31	75.6	2342	5	001677	001677 bombay mori
37	30	73.2	96	9	048440	048440 bacterioph
38	30	73.2	129	5	062211	062211 caenorhabd1
39	30	73.2	141	5	096254	096254 plasmodium
40	30	73.2	153	14	093567	093567 fowlpox vir
41	30	73.2	163	5	061832	061832 caenorhabd1
42	30	73.2	168	2	031938	031938 bacillus su
43	30	73.2	168	9	064083	064083 bacterioph
44	30	73.2	254	11	062891	062891 rattus norv
45	30	73.2	267	2	091060	091060 pseudomonas

ALIGNMENTS

RESULT	1	PRELIMINARY	PRT	171 AA
Q9WUW5	Q9WUW5			
AC	Q9WUW5			
DT	01-NOV-1999 (TREMBLrel. 12, Created)			
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)			
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)			
DE	COLLAGEN TYPE XVIII, ALPHA (1) CHAIN (FRAGMENT).			
GN	COL18A1.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LIVER.			
RA	Jia J.D., Bauer M., Ederpacher U., Donner P., Schuppan D.;			
RT	"Partial 3'-terminal cDNA sequence of rat collagen XVIII/endostatin."			
RL	Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AJ236873; CAB44263.1; -.			
DR	HSSP: P39061; IKOE.			
KW	Collagen.			
FT	NON_TER			
FT	NON_TER			
SO	SEQUENCE			

Query Match 100.0%; Score 41; DB 11; Length 171;
Best Local Similarity 100.0%; Pred. No. 0.36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
DB 160 SYIVLCIE 167

```

RESULT 2
Q9JK63 PRELIMINARY: PRT: 184 AA.
AC Q9JK63;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE ENDOSTATIN (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHINESE KUNMING;
RA Jia S., Zhu F., Xing G., Yu Y., Duan C., Xiu R.-J., He F.;
RT "Anticancer treatment of targeted fusion protein delivery to tumor
RT neovasculature."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF257775; AAF69009.1; -.
DR NON_TER 1
FT NON_TER 184
SQ SEQUENCE 184 AA: 20376 MW: AC06F9D8D103412A CRC64;

```

```

Query Match 100.0%; Score 41; DB 11; Length 184;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 SYVLICIE 8
Db 168 SYVLICIE 175

```

```

RESULT 3
Q9QZD2 PRELIMINARY: PRT: 226 AA.
AC Q9QZD2;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE COLLAGEN XVIII (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RA Chen L., Perletti G., Folkman J.;
RT "Antitumor activity of rat endostatin."
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF189709; AAF00975.1; -.
DR HSSP: P39061; IKOE.
KM collagen.
RT chain.
FT NON_TER 1
SQ SEQUENCE 226 AA: 25350 MW: 38B83C0486C0E949 CRC64;

```

```

Query Match 100.0%; Score 41; DB 11; Length 226;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 SYVLICIE 8
Db 210 SYVLICIE 217

```

```

RESULT 4
O61434 PRELIMINARY: PRT: 1140 AA.
AC O61434;

```

```

DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update).
DE 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE COLLAGEN (FRAGMENT).
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94059075; Pubmed=8240330;
RA Abe N., Muragaki Y., Yoshida K., Inoue H., Ninomiya Y.;
RT Identification of a novel collagen chain represented by extensive
RT interruptions in the triple-helical region.
RL Cell. Mol. Biol. Res. 196; 576-582(1993).
DR EMBL: D17546; BAA04483.1; -.
DR HSSP: P39061; IKOE.
DR MGD: MGI:88449; COL15A1.
DR InterPro: IPR000087; -.
DR Pfam: PF01391; Collagen; 6.
FT NON_TER 1
SQ SEQUENCE 1140 AA: 115156 MW: 8B0C7E6862B3BDFE CRC64;

```

```

Query Match 100.0%; Score 41; DB 11; Length 1140;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 SYVLICIE 8
Db 1124 SYVLICIE 1131

```

```

RESULT 5
Q62001 PRELIMINARY: PRT: 1774 AA.
AC Q62001; Q60672;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE PROCOLLAGEN, TYPE XVIII, ALPHA 1 PRECURSOR (XVIII) COLLAGEN
DE (PROCOLLAGEN, TYPE XVIII, ALPHA 1) (ALPHA-1 TYPE XVIII COLLAGEN).
GN COL18A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PRIMARY TAIL CULTURE;
RX MEDLINE=94245707; Pubmed=8188673;
RA Rehn M., Hinkka E., Pihlajaniemi T.;
RT Primary structure of the alpha 1 chain of mouse type XVIII collagen,
RT partial structure of the corresponding gene, and comparison of the
RT alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen
RT chain.
RL J. Biol. Chem. 269; 13929-13935(1994).
RN [2]
RP SEQUENCE OF 1-562 FROM N.A.
RX MEDLINE=94240112; Pubmed=8183894;
RA Rehn M., Pihlajaniemi T.;
RT Alpha 1(XVIII), a collagen chain with frequent interruptions in the
RT collagenous sequence, a distinct tissue distribution, and homology
RT with type XV collagen.
RL Proc. Natl. Acad. Sci. U.S.A. 91; 4234-4238(1994).
RN [3]
RP SEQUENCE OF 1-562 FROM N.A.
RX MEDLINE=95181468; Pubmed=7876242;
RA Rehn M., Pihlajaniemi T.;
RT Identification of three N-terminal ends of type XVIII collagen chains
RT and tissue-specific differences in the expression of the corresponding
RT transcripts. The longest form contains a novel motif homologous to rat

```

RF and Drosophila frizzled proteins.";
 RL J. Biol. Chem. 270:4705-4711(1995).
 DR EMBL: U03715; AAC52903.1; -.
 DR EMBL: U03716; AAC52903.1; JOINED.
 DR EMBL: U03718; AAC52903.1; JOINED.
 DR EMBL: U34607; AAC52903.1; JOINED.
 DR EMBL: U34608; AAC52903.1; JOINED.
 DR EMBL: U34609; AAC52903.1; JOINED.
 DR EMBL: U34610; AAC52903.1; JOINED.
 DR EMBL: U34611; AAC52903.1; JOINED.
 DR EMBL: U34612; AAC52903.1; JOINED.
 DR EMBL: U34613; AAC52903.1; JOINED.
 DR EMBL: U11637; AAC52179.1; -.
 DR HSSP: P39061; IKOE.
 DR MGD: MGI:88451; Coll18a1.
 DR InterPro: IPR000024; -.
 DR InterPro: IPR001087; -.
 DR InterPro: IPR001791; -.
 DR InterPro: IPR003129; -.
 DR Pfam: PF01391; Collagen; 6.
 DR Pfam: PF01392; Fz; 1.
 DR Pfam: PF02210; TSPN; 1.
 DR SMART: SM00282; Lang; 1.
 KW Signal.
 SQ SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88EF232 CRC64;

Query Match 100.0%; Score 41; DB 11; Length 1774;
 Best Local Similarity 100.0%; Pred. No. 2.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SYVLCIE 8
 |||||
 Db 1758 SYVLCIE 1765

RESULT 6
 ID 087543 PRELIMINARY; PRT: 237 AA.
 AC 087543;
 DT 01-NOV-1998 (TEMBLrel. 08, Created)
 DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
 DT 01-OCT-2000 (TEMBLrel. 15, Last annotation update)
 DE OUTER MEMBRANE PROTEIN ADHESIN (FRAGMENT).
 OS Capnocytophaga gingivalis.
 OC Bacteria; CPB group; Flavobacteriia; Flavobacteriaceae; Capnocytophaga.
 OX NCBI_TaxID=1017;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DR2001;
 RA Tempo P.J.;
 RT "Capnocytophaga outer-membrane adhesin DNA sequence, 3'end.";
 RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF083593; AAC35353.1; -.
 DR MEROPS: S09.013; -.
 DR InterPro: IPR000379; -.
 DR InterPro: IPR001375; -.
 DR InterPro: IPR002471; -.
 DR Pfam: PF00326; Peptidase_S9; 1.
 DR PROSITE: PS00708; PRO_ENDOPPEP_SER; 1.
 FT NON_TER
 SQ SEQUENCE 237 AA; 27482 MW; 4D630AC81D64EDAC CRC64;

Query Match 80.5%; Score 33; DB 2; Length 237;
 Best Local Similarity 71.4%; Pred. No. 23;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 2 YIVLCIE 8
 |||||
 Db 62 YIVLCVD 68

RESULT 7
 ID 093419 PRELIMINARY; PRT: 386 AA.
 AC 093419;
 DT 01-NOV-1998 (TEMBLrel. 08, Created)
 DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
 DT 01-JUN-2000 (TEMBLrel. 14, Last annotation update)
 DE COLLAGEN XVIII (FRAGMENT).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauia; Metazoa; Chordata; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Halfter W., Dong S., Schurer B., Cole G.;
 RT "Collagen XVIII is a basement membrane heparan sulfate proteoglycan.";
 RL J. Biol. Chem. 0:0-0(1998).
 DR EMBL: AF083440; AAC33294.1; -.
 DR HSSP: P39061; IKOE.
 DR InterPro: IPR000087; -.
 FT NON_TER
 SQ SEQUENCE 386 AA; 41775 MW; 34DA0FA09EBA3B0E CRC64;

Query Match 80.5%; Score 33; DB 13; Length 386;
 Best Local Similarity 62.5%; Pred. No. 35;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SYVLCIE 8
 ::|||
 Db 370 AFVLCIE 377

RESULT 8
 ID 015900 PRELIMINARY; PRT: 646 AA.
 AC 015900;
 DT 01-JAN-1998 (TEMBLrel. 05, Created)
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
 DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
 DE PUTATIVE TRANSCRIPTION INHIBITOR.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellum.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-TDD-4;
 RA Wells D.J., Welker D.L.;
 RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U57081; AB95435.1; -.
 DR InterPro: IPR001584; -.
 DR Pfam: PF00665; Iwe; 1.
 SQ SEQUENCE 646 AA; 74364 MW; 89EAF091505B2D4E CRC64;

Query Match 80.5%; Score 33; DB 5; Length 646;
 Best Local Similarity 71.4%; Pred. No. 55;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 2 YIVLCIE 8
 |||||
 Db 185 YIILCID 191

RESULT 9
 ID 015901 PRELIMINARY; PRT: 706 AA.
 AC 015901;
 DT 01-JAN-1998 (TEMBLrel. 05, Created)
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
 DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
 DE PUTATIVE TRANSPORASE.

OS Dictyostellum discoideum (slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostellum.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TRANSDON-TD-4:
 RA Wells D.J., Melker D.L.:
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: 057081; AAB95436.1; -;
 DR InterPro: IPR001584; -;
 DR Pfam: PF00665; rve; 1.
 SQ SEQUENCE 706 AA; 80670 MW; F5718916484A3E5 CRC64;

Query Match 80.5%; Score 33; DB 5; Length 706;
 Best Local Similarity 71.4%; Pred. No. 59;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 SYVLCT 8
 DB 185 SYLCTD 191

RESULT 10
 ID 09XTP6 PRELIMINARY; PRT; 1905 AA.
 AC 09XTP6;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE CTRP PROTEIN PRECURSOR (OOKINETE PROTEIN).
 OS Plasmodium berghei.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5821;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ANKA;
 RA Yuda M., Sawai T., Chinzel Y.;
 RL "Structure and expression of an adhesive protein-like molecule of mosquito invasive-stage malaria parasite."
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ANKA;
 RA Yuda M., Sawai T., Chinzel Y.;
 RL "Structure and expression of an adhesive protein-like molecule of mosquito invasive-stage malaria parasite."
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ANKA;
 RA Yuda M.;
 RL "Structure and Expression of an Adhesive Protein-Like Molecule of Mosquito Invasive-stage Malarial Parasite.";

RT J. Exp. Med. 0:0-0(1999)
 RL EMBL: A238798; CAB45562.1; -;
 DR EMBL: A5027129; BA8322.1; -;
 DR EMBL: AF149771; AAF73158.1; -;
 DR InterPro: IPR000884; -;
 DR InterPro: IPR002035; -;
 DR Pfam: PF00090; tsg_1; 4.
 DR Pfam: PF00092; vva; 5.
 DR PRINTS: PRO0453; VWFADOMAIN.
 KW SIGNAL.
 FT CHAIN 1 21 POTENTIAL.
 DR SIGNAL 22 1905 CTRP PROTEIN.
 SQ SEQUENCE 1905 AA; 215138 MW; 27A94B778CABDC36 CRC64;

Query Match 80.5%; Score 33; DB 5; Length 1905;
 Best Local Similarity 71.4%; Pred. No. 1.4e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYVLCT 7
 DB 10 SYVLCT 16

RESULT 11
 ID 084569 PRELIMINARY; PRT; 147 AA.
 AC 084569;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE HYPOTHETICAL 16.0 KDA PROTEIN.
 GN CT565.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-D/OW-3/CX;
 RX MEDLINE-99000809; PubMed=9784136;
 RA Stephens R.S., Kalman S., Lamm C.J., Fan J., Marathe R., Aravind L., Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., Davis R.W.;
 RA "Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis."
 RL Science 282:754-759(1998).
 DR EMBL: AE001327; AAC68167.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 147 AA; 15980 MW; 19D0E2CE2F9DE4E CRC64;

Query Match 78.0%; Score 32; DB 2; Length 147;
 Best Local Similarity 85.7%; Pred. No. 26;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYVLCT 7
 DB 123 SYVLCT 129

RESULT 12
 ID 09PJH5 PRELIMINARY; PRT; 147 AA.
 AC 09PJH5;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE HYPOTHETICAL PROTEIN TC0854.
 GN TC0854.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MOPN / NIGS;
 RX MEDLINE-20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Koloney J., McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
 RA "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39."
 RT Nucleic Acids Res. 28:1397-1406(2000).
 RL EMBL: AE003352; AAF39650.1; -;
 DR TIGR: TC0854; -;
 KW Hypothetical protein.
 SQ SEQUENCE 147 AA; 15871 MW; 6B33E9C475F8FD66 CRC64;

Query Match 78.0%; Score 32; DB 2; Length 147;

Best Local Similarity 85.7%; Pred. No. 26;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCI 7
|:|||||
Db 123 SYIVLCI 129

RESULT 13
ID 09P9G6 PRELIMINARY; PRT; 347 AA.
AC 09P9G6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HYPOHETICAL 39.0 KDA PROTEIN.
OS Methanococcoides burtonii.
OC Archaea; Euryarchaeota; Methanosarcinales; Methanosarcinaceae;
OC Methanococcoides.
OX NCBI_TaxID=29291;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 6242;
RX MEDLINE=20198308; PubMed=10731411;
RA Lim J., Thomas T., Cavichioni R.;
RT "Low temperature regulated DEAD-box RNA helicase from the Antarctic
archaeon, Methanococcoides burtonii.";
RL J. Mol. Biol. 297:553-567(2000).
DR EMBL: AF199442; AAF89100.1; -;
KW Hypothetical protein.
SQ SEQUENCE 347 AA; 39010 MW; EFP2AC2CDIF6DC8 CRC64;

Query Match 78.0%; Score 32; DB 1; Length 347;
Best Local Similarity 85.7%; Pred. No. 53;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCI 7
|:|||||
Db 196 SYIVLCI 202

RESULT 14
ID 09P905 PRELIMINARY; PRT; 507 AA.
AC 09P905;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE PALC.
GN PALC.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eutriciales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RA Negrete-Urtasun S., Reiter W., Diez E., Denison S.H., Tilburn J.,
RA Espeso E.A., Penava M.A., Arst H.N. Jr.;
RT "Ambient pH signal transduction in Aspergillus: completion of gene
RT characterization.";
RL Mol. Microbiol. 0:0-0(1999).
DR EMBL: AF152925; AAF70857.1; -;
SQ SEQUENCE 507 AA; 54580 MW; 3AA315413977F929 CRC64;

Query Match 78.0%; Score 32; DB 3; Length 507;
Best Local Similarity 62.5%; Pred. No. 72;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYIVLCI 8
|:|||||
Db 236 SYIVLCI 243

RESULT 15
ID 095942 PRELIMINARY; PRT; 613 AA.
AC 095942;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE ALPHA INTEGRIN INTERACTING PROTEIN 63 (FRAGMENT).
GN AIBP63.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Wixler V.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-235 FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=99192311; PubMed=10094488;
RA Wixler V., Laplantine E., Geerts D., Sonnenberg A., Petersohn D.,
RA Ecker B., Paulsson M., Ammalle M.;
RT "Identification of novel interaction partners for the conserved
RT membrane proximal region of alpha-integrin cytoplasmic domains.";
RL PNAS Lett. 445:351-355(1999).
DR EMBL: AJ131721; CAB38232.1; -;
KW Integrin.
FT NON_TER
SQ SEQUENCE 613 AA; 69523 MW; F742491675A63F55 CRC64;

Query Match 78.0%; Score 32; DB 4; Length 613;
Best Local Similarity 57.1%; Pred. No. 85;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 YIVLCIE 8
|:|||||
Db 473 YIVLCIE 479

Search completed: August 16, 2001, 11:44:01
Job time: 174 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 11:41:28 ; Search time 10.05 Seconds
(without alignments)
27.268 Million cell updates/sec

Title: US-09-589-777a-2_COPY_168_175

Perfect score: 41
Sequence: 1 SYVLICIE 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	1527	1 CAIH_MOUSE	P39061 mus musculu
2	38	92.7	1516	1 CAIH_HUMAN	P39060 homo sapien
3	32	78.0	1015	1 YAI7_HUMAN	O9unp23 homo sapien
4	30	73.2	233	1 YAS6_METUA	O58442 methanococc
5	30	73.2	346	1 RDS_FEICA	P35306 felis silve
6	30	73.2	407	1 REC_SALTY	P26479 salmoneila
7	30	73.2	420	1 NAPI_HUMAN	O96009 homo sapien
8	30	73.2	436	1 YW40_YEAST	O00624 homo sapien
9	30	73.2	550	1 COMP_BAC2U	O03212 saccharomyc
10	30	73.2	769	1 TH11_SCHPO	O99027 bacillus su
11	30	73.2	775	1 VP11_NEUCR	P34558 schizosacch
12	30	73.2	856	1 CALE_HUMAN	O01280 neurospora
13	30	73.2	880	1 YOTO_CAEBL	P39059 homo sapien
14	29	70.7	1388	1 MCSR_HUMAN	P33032 caenorhabdi
15	29	70.7	325	1 MOP4_HUMAN	O15374 homo sapien
16	29	70.7	467	1 CRK_DAUCA	P33681 daucus caro
17	29	70.7	891	1 YB31_SCHPO	O14338 schizosacch
18	29	70.7	923	1 YAU3_SCHPO	O10165 schizosacch
19	29	70.7	933	1 LRP1_CHICK	P08157 gallus gall
20	29	70.7	4543	1 PEPB_PIG	P15782 sus scrofa
21	28	68.3	79	1 YN25_CAEBL	P31584 caenorhabdi
22	28	68.3	125	1 HBB_MEGLY	P11575 megaderma 1
23	28	68.3	146	1 Y173_METUA	O57637 methanococc
24	28	68.3	157	1 IRI0_HCTVA	P16808 human cytom
25	28	68.3	171	1 LACB_PIG	P04119 sus scrofa
26	28	68.3	178	1 AAT2_VACCV	P24758 vaccinia vi
27	28	68.3	213	1 VHEL_MCMVM	P00499 white clove
28	28	68.3	236	1 VHEL_MCMVM	P14403 bottle clove
29	28	68.3	253	1 TPIS_BORBU	O51982 borrelia bu
30	28	68.3	255	1 YFRC_NEIMB	P32015 neisseria pe
31	28	68.3	294	1 YFRC_YERPE	O56954 yersinia pe
32	28	68.3	305	1 CAG7_RAT	O64686 rattus norv
33	28	68.3	305	1 CAG7_RAT	O64686 rattus norv

34	28	68.3	322	1 AAT2_VACCC	P21114 vaccinia v1
35	28	68.3	325	1 MCSR_BOVIN	P56451 bos taurus
36	28	68.3	340	1 YDDR_ECOLI	P77308 escherichia
37	28	68.3	353	1 GPRO_RAT	P97639 rattus norv
38	28	68.3	381	1 YCU4_SCHPO	O74405 schizosacch
39	28	68.3	402	1 GPRO_HUMAN	O99705 homo sapien
40	28	68.3	435	1 YU49_HSVSA	O01046 herpesvirus
41	28	68.3	496	1 YKAB_CAEBL	P34262 caenorhabdi
42	28	68.3	525	1 SYH_CAEBL	P34183 caenorhabdi
43	28	68.3	540	1 YFEO_YEAST	P43562 saccharomyc
44	28	68.3	560	1 FOBI_YEAST	O13329 saccharomyc
45	28	68.3	596	1 ACP2_YEAST	P38090 saccharomyc

ALIGNMENTS

RESULT 1
CAIH_MOUSE STANDARD: PRT: 1527 AA.
ID CAIH_MOUSE Q62002, Q61437.
AC P39061; Q62002, Q61437.
DT 01-FEB-1995 (Rel. 31, Created)
DR 01-OCT-2000 (Rel. 40, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE COLLAGEN ALPHA 1(XVIII) CHAIN PRECURSOR [CONSTAINS: ENDOSTATIN].
GN COL18A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid:10090;
RN [1]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RC STRAIN-BALB/C; TISSUE-Liver;
RX MEDLINE-94245707; PubMed-8188673;
RA Rein M.V., Hintsikka E., Pihlajaniemi T.;
RT "Primary structure of the alpha 1 chain of mouse type XVIII collagen, partial structure of the corresponding gene, and comparison of the alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen chain.";
RT J. Biol. Chem. 269:13929-13935(1994).
RL [2]
RN SEQUENCE FROM N.A. (SHORT AND LONG ISOFORMS).
RP Rein M., Hintsikka E., Pihlajaniemi T.;
RX "Characterization of the mouse gene for the alpha-1 chain of type XVIII collagen (COL18A1) reveals that the three variant N-terminal polypeptide forms are transcribed from two widely separated promoters.";
RT Submitted (NOV-1993) to the EMBL/Genbank/DBJ databases.
RL [3]
RN SEQUENCE OF 213-1140 FROM N.A. (SHORT ISOFORM).
RP MEDLINE-94240112; PubMed-8183894;
RX Rein M.V., Pihlajaniemi T.;
RA "Alpha 1(XVIII), a collagen chain with frequent interruptions in the collagenous sequence, a distinct tissue distribution, and homology with type XV collagen.";
RT Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).
RL [4]
RN SEQUENCE OF 240-1527 FROM N.A.
RP TISSUE-Liver;
RX MEDLINE-94240111; PubMed-8183893;
RA Oh S.P., Kamagata Y., Muragaki Y., Jimmons S., Ooshima A., Olsen B.R.;
RT "Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa-Yaa repeats identify a distinct family of collagenous proteins.";
RT Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).
RL [5]
RN CHARACTERIZATION OF ENDOSTATIN AND PARTIAL SEQUENCE.
RP MEDLINE-94240118; PubMed-9008168;
RX O'Reilly M.S., Boehm T., Shing Y., Fukai N., Vasios G., Lane W.S., Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;
RT "Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.";
RT Cell 88:277-285(1997).

bioinformatics.sib.ch. All rights of reproduction or entry are copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on the use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; L16898; AAA37434.1; -

```
ENBL; U03714; AAA20657.1; -.
EMBL; U03715; AAC52063.1; -.

```

EMBL; U34606; AAC52901.1; -; JOINED

EMBL; U34608; AAC52901.1; JOINED.

```
EMBL: U34609; AAC52901.1; JOINED.  
EMBL: U34610; AAC52901.1; JOINED.
```

EMBL; U34611; AAC52901.1; JOINED.

EMBL; U34612; AAC52901.1; JOINED.
EMBL; U34613; AAC52902.1; JOINED.
EMBL; U34614; AAC52903.1; JOINED.

EMBL: U03716; AAC52901.1; JOINED.
EMBL: U03716; AAC52901.1; JOINED.

EMBL; U03718; AAC52901.1; JOINED.

EMBL: U03715; AAC52902.1; -
EMBL: U34507; AAC50000.1; -

EMBL; U34608; AAC52902.1; JOINED.

EMBL; U34609; AAC52902.1; JOINED.

EMBL: U34610; AAC52902.1; JOINED.
EMBL: U34611; AAC52903.1; JOINED.

EMBL; U34612; AAC52902.1; JOINED.

EMBL; U34613; AAC52902.1; JOINED.
MBI : F03775
MBI : F03775

```

MBL; 003718; AAC52902.1; JOINED.

```

```

      MBL; U11636; AAC52178.1; - COINED.

```

```

MBL; L22545; AAA19787.1; -
DR: 1KOE: 16-FEB-00

```

GD; MGI:88451; C0118a1

```

interPro; IPR000087; .

```

ram; PF01391; Collagen; 6.

cell adhesion; Collagen; Connective tissue; Repeat; Hydroxylation;

D-structure.

- cytochrome; signal; Alternative splicing;

CHAIN	POTENTIAL.
1	26
27	1527

CHAIN	M_n	M_w	M_w/M_n
COLLAGEN ALPHA 1(XVII) CHAIN.	1527	1527	1.00
ENDOSTATIN	1344	1527	1.14

NONHELIICAL REGION 1 (NC1)	
538	END OF MAIN
27	END OF MAIN

Region	Residue	Structure
TRIPLE-HELICAL REGION 1 (COL1):	565	565
	566	566
	567	567
TRIPLE-HELICAL REGION 2 (COL1):	568	568
	569	569
	570	570
TRIPLE-HELICAL REGION 3 (COL1):	571	571
	572	572
	573	573
TRIPLE-HELICAL REGION 4 (COL1):	574	574
	575	575
	576	576
TRIPLE-HELICAL REGION 5 (COL1):	577	577
	578	578
	579	579
TRIPLE-HELICAL REGION 6 (COL1):	580	580
	581	581
	582	582
TRIPLE-HELICAL REGION 7 (COL1):	583	583
	584	584
	585	585
TRIPLE-HELICAL REGION 8 (COL1):	586	586
	587	587
	588	588
TRIPLE-HELICAL REGION 9 (COL1):	589	589
	590	590
	591	591
TRIPLE-HELICAL REGION 10 (COL1):	592	592
	593	593
	594	594
TRIPLE-HELICAL REGION 11 (COL1):	595	595
	596	596
	597	597
TRIPLE-HELICAL REGION 12 (COL1):	598	598
	599	599
	600	600
TRIPLE-HELICAL REGION 13 (COL1):	601	601
	602	602
	603	603
TRIPLE-HELICAL REGION 14 (COL1):	604	604
	605	605
	606	606
TRIPLE-HELICAL REGION 15 (COL1):	607	607
	608	608
	609	609
TRIPLE-HELICAL REGION 16 (COL1):	610	610
	611	611
	612	612
TRIPLE-HELICAL REGION 17 (COL1):	613	613
	614	614
	615	615
TRIPLE-HELICAL REGION 18 (COL1):	616	616
	617	617
	618	618
TRIPLE-HELICAL REGION 19 (COL1):	619	619
	620	620
	621	621
TRIPLE-HELICAL REGION 20 (COL1):	622	622
	623	623
	624	624
TRIPLE-HELICAL REGION 21 (COL1):	625	625
	626	626
	627	627
TRIPLE-HELICAL REGION 22 (COL1):	628	628
	629	629
	630	630
TRIPLE-HELICAL REGION 23 (COL1):	631	631
	632	632
	633	633
TRIPLE-HELICAL REGION 24 (COL1):	634	634
	635	635
	636	636
TRIPLE-HELICAL REGION 25 (COL1):	637	637
	638	638
	639	639
TRIPLE-HELICAL REGION 26 (COL1):	640	640
	641	641
	642	642
TRIPLE-HELICAL REGION 27 (COL1):	643	643
	644	644
	645	645
TRIPLE-HELICAL REGION 28 (COL1):	646	646
	647	647
	648	648
TRIPLE-HELICAL REGION 29 (COL1):	649	649
	650	650
	651	651
TRIPLE-HELICAL REGION 30 (COL1):	652	652
	653	653
	654	654
TRIPLE-HELICAL REGION 31 (COL1):	655	655
	656	656
	657	657
TRIPLE-HELICAL REGION 32 (COL1):	658	658
	659	659
	660	660
TRIPLE-HELICAL REGION 33 (COL1):	661	661
	662	662
	663	663
TRIPLE-HELICAL REGION 34 (COL1):	664	664
	665	665
	666	666
TRIPLE-HELICAL REGION 35 (COL1):	667	667
	668	668
	669	669
TRIPLE-HELICAL REGION 36 (COL1):	670	670
	671	671
	672	672
TRIPLE-HELICAL REGION 37 (COL1):	673	673
	674	674
	675	675
TRIPLE-HELICAL REGION 38 (COL1):	676	676
	677	677
	678	678
TRIPLE-HELICAL REGION 39 (COL1):	679	679
	680	680
	681	681
TRIPLE-HELICAL REGION 40 (COL1):	682	682
	683	683
	684	684
TRIPLE-HELICAL REGION 41 (COL1):	685	685
	686	686
	687	687
TRIPLE-HELICAL REGION 42 (COL1):	688	688
	689	689
	690	

DOMAIN	576	649	713
NONHELICAL REGION 2 (NC2).			
TRIPLE-HELICAL REGION 2			

	NONHELICAL REGION 2 (COL2).	NONHELICAL REGION 3 (NC3)
650		
673		
674		
675		
676		
677		
678		
679		
680		
681		
682		
683		
684		
685		
686		
687		
688		
689		
690		
691		
692		
693		
694		
695		
696		
697		
698		
699		
700		
701		
702		
703		
704		
705		
706		
707		
708		
709		
710		
711		
712		
713		
714		
715		
716		
717		
718		
719		
720		
721		
722		
723		
724		
725		
726		
727		
728		
729		
730		
731		
732		
733		
734		
735		
736		
737		
738		
739		
740		
741		
742		
743		
744		
745		
746		
747		
748		
749		
750		
751		
752		
753		
754		
755		
756		
757		
758		
759		
760		
761		
762		
763		
764		
765		
766		
767		
768		
769		
770		
771		
772		
773		
774		
775		
776		
777		
778		
779		
780		
781		
782		
783		
784		
785		
786		
787		
788		
789		
790		
791		
792		
793		
794		
795		
796		
797		
798		
799		
800		
801		
802		
803		
804		
805		
806		
807		
808		
809		
810		
811		
812		
813		
814		
815		
816		
817		
818		
819		
820		
821		
822		
823		
824		
825		
826		
827		

	795	TRIPLE-HELICAL REGION 3 (COL3).
6/4	796	
MAIN	818	

	NONHELICAL REGION 4 (NC4).
819	TRIPLE-HELICAL REGION 4.
901	
MAIN	

	902	915	922
MAIN			
NONHELICAL REGION 4 (COL4)			
NONHELICAL REGION 5 (NC5)			
MAIN			

916
958
957
970

TRIPLE-HELICAL REGION 5 (COL5) -
MOVING TARGET

MAIN	971	1043	NONHELICAL REGION 6 (NC6); TRIPLE-HELICAL REGION 6 (NC6)
220			

MAIN	1044	1053	NONHELICAL REGION 6 (COL6).
MAIN	1054	1055	NONHELICAL REGION 7 (NC7).

1034 1086 TRIPLE-HELICAL REGION 7 (COL7).

FT	DOMAIN	1087	1098	1098		NONHELICAL REGION 8 (NC8).
FT	DOMAIN	1089	1122	1122		TRIPLE-HELICAL REGION 8 (COL8).
FT	DOMAIN	1123	1129	1129		NONHELICAL REGION 9 (NC9).
FT	DOMAIN	1130	1181	1181		TRIPLE-HELICAL REGION 9 (COL9).
FT	DOMAIN	1182	1194	1194		NONHELICAL REGION 10 (NC10).
FT	DOMAIN	1195	1212	1212		TRIPLE-HELICAL REGION 10 (COL10).
FT	DOMAIN	1213	1527	1527		NONHELICAL REGION 11 (NC11).
FT	CARBOHYD	338	338	338		N-LINKED (GLCNAC..) (POTENTIAL).
FT	CARBOHYD	700	700	700		N-LINKED (GLCNAC..) (POTENTIAL).
FT	DISULFID	1376	1516	1516		N-LINKED (GLCNAC..) (POTENTIAL).
FT	DISULFID	1478	1508	1508		N-LINKED (GLCNAC..) (POTENTIAL).
FT	SITE	1104	1106	1106		
FT	VARSPLIC	1	212	212		CELL ATTACHMENT SITE (POTENTIAL).
FT	VARSPLIC	213	238	238		MISSING (IN SHORT ISOFORM).
FT	CONFLICT	900	900	900		APPTDPPGNSLOAIPGRAPPDF -> MAPRHHLLDVLP TSVLIVLVARVSME (IN SHORT ISOFORM).
FT	CONFLICT	947	947	947		P -> L (IN REF. 4).
FT	CONFLICT	964	964	964		A -> F (IN REF. 4).
FT	CONFLICT	1157	1157	1157		R -> R (IN REF. 4).
FT	CONFLICT	1266	1266	1266		P -> P (IN REF. 4).
FT	CONFLICT	1276	1276	1276		L -> F (IN REF. 4).
FT	CONFLICT	1437	1437	1437		L -> V (IN REF. 4).
SEQ	SEQUENCE	1527	AA: 156008	MW: 96455045AET140E513	CPCE64.	

```

Query Match      100.0%; Score 41; DB 1; Length 1527;
Best Local Similarity 100.0%; Pred. No. 1,7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
2Y      1 SYIVLCIE 8
          |||||
db      1511 SYIVCIE 1518

```

Accession	Gene	Species	Standard	Protein	AA
U00001	COL1A1	HUMAN			
U00002	COL1A2	HUMAN			
U00003	COL1A3	HUMAN			
U00004	COL1A4	HUMAN			
U00005	COL1A5	HUMAN			
U00006	COL1A6	HUMAN			
U00007	COL1A7	HUMAN			
U00008	COL1A8	HUMAN			
U00009	COL1A9	HUMAN			
U00010	COL1A10	HUMAN			
U00011	COL1A11	HUMAN			
U00012	COL1A12	HUMAN			
U00013	COL1A13	HUMAN			
U00014	COL1A14	HUMAN			
U00015	COL1A15	HUMAN			
U00016	COL1A16	HUMAN			
U00017	COL1A17	HUMAN			
U00018	COL1A18	HUMAN			
U00019	COL1A19	HUMAN			
U00020	COL1A20	HUMAN			
U00021	COL1A21	HUMAN			
U00022	COL1A22	HUMAN			
U00023	COL1A23	HUMAN			
U00024	COL1A24	HUMAN			
U00025	COL1A25	HUMAN			
U00026	COL1A26	HUMAN			
U00027	COL1A27	HUMAN			
U00028	COL1A28	HUMAN			
U00029	COL1A29	HUMAN			
U00030	COL1A30	HUMAN			
U00031	COL1A31	HUMAN			
U00032	COL1A32	HUMAN			
U00033	COL1A33	HUMAN			
U00034	COL1A34	HUMAN			
U00035	COL1A35	HUMAN			
U00036	COL1A36	HUMAN			
U00037	COL1A37	HUMAN			
U00038	COL1A38	HUMAN			
U00039	COL1A39	HUMAN			
U00040	COL1A40	HUMAN			
U00041	COL1A41	HUMAN			
U00042	COL1A42	HUMAN			
U00043	COL1A43	HUMAN			
U00044	COL1A44	HUMAN			
U00045	COL1A45	HUMAN			
U00046	COL1A46	HUMAN			
U00047	COL1A47	HUMAN			
U00048	COL1A48	HUMAN			
U00049	COL1A49	HUMAN			
U00050	COL1A50	HUMAN			
U00051	COL1A51	HUMAN			
U00052	COL1A52	HUMAN			
U00053	COL1A53	HUMAN			
U00054	COL1A54	HUMAN			
U00055	COL1A55	HUMAN			
U00056	COL1A56	HUMAN			
U00057	COL1A57	HUMAN			
U00058	COL1A58	HUMAN			
U00059	COL1A59	HUMAN			
U00060	COL1A60	HUMAN			
U00061	COL1A61	HUMAN			
U00062	COL1A62	HUMAN			
U00063	COL1A63	HUMAN			
U00064	COL1A64	HUMAN			
U00065	COL1A65	HUMAN			
U00066	COL1A66	HUMAN			
U00067	COL1A67	HUMAN			
U00068	COL1A68	HUMAN			
U00069	COL1A69	HUMAN			
U00070	COL1A70	HUMAN			
U00071	COL1A71	HUMAN			
U00072	COL1A72	HUMAN			
U00073	COL1A73	HUMAN			
U00074	COL1A74	HUMAN			
U00075	COL1A75	HUMAN			
U00076	COL1A76	HUMAN			
U00077	COL1A77	HUMAN			
U00078	COL1A78	HUMAN			

SEQUENCE FROM N.A.
MEDLINE-20289799; PubMed-10830953;
Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
Park H.-S., Toyoda A., Ishii K., Teroaki Y., Choi D.-K., Sueda E.,
Ohni M., Takagi T., Sakaki Y., Tauden S., Blechschmidt K., Polley A.,
Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
Roenthal A., Kidon J., Shibuya K., Kawasaki K., Asakawa S.,
Shinaiti A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
Minoshima S., Shimizu N., Nordstok G., Hornischer K., Brindt P.,
Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
Ramser J., Beck A., Klaes S., Hennig S., Resselmann L., Dagnan E.,
Wehmeyer S., Borym K., Gardiner K., Nizetic D., Francis F.,
Leirich H., Reinhardt R., Yaspo M.-L.,
The DNA sequence of human chromosome 21.
Nature 405:311-319(2000).
[31]

RP SEQUENCE OF 834-1516 FROM N.A.
 RX MEDLINE-94245237; PubMed-8188291;
 RA Oh S.P., Wairman M.L., Seldin M.F., Cheng S., Knoll J.H., Timmons S.,
 RA Olsen B.R.;
 RT *Cloning of cDNA and genomic DNA encoding human type XVIII collagen
 RT and localization of the alpha 1(XVIII) collagen gene to mouse
 RT chromosome 10 and human chromosome 21.*;
 RL Genomics 19:494-499(1994).
 RN [4]
 RP SEQUENCE OF 1334-1516 FROM N.A.
 RC TISSUE-Placenta;
 RA Zhi-Yong H., Biao L., Wei-Jie Z., Xiang-Fu W.;
 RT *Cloning and expression of human endostatin gene in Escherichia
 RT coli.*;
 RL submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP INVOLVEMENT IN KNOBLOCH SYNDROME.
 RX MEDLINE-20400145; PubMed-10942434;
 RA Sertle A.L., Sossi V., Camargo A.A., Zatz M., Brahe C.,
 RA Passos-Bueno M.R.;
 RT *Collagen XVIII, containing an endogenous inhibitor of angiogenesis
 RT and tumor growth, plays a critical role in the maintenance of retinal
 RT structure and in neural tube closure.*;
 RL Hum. Mol. Genet. 9:2051-2058(2000)
 CC -1- FUNCTION: COL1A18A PROBABLY PLAYS A MAJOR ROLE IN DETERMINING THE
 CC RETINAL STRUCTURE AS WELL AS IN THE CLOSURE OF THE NEURAL TUBE.
 CC -1- FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL
 CC PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY
 CC BINDING TO THE HEPARAN SULFATE PROTEOGLYCANS INVOLVED IN GROWTH
 CC FACTOR SIGNALING (BY SIMILARITY).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOMORFS: A SHORT FORM/NC1-103 AND A LONG
 CC FORM/NC-493 (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING
 CC -1- TISSUE SPECIFICITY: PRESENT IN MULTIPLE ORGANS WITH HIGHEST LEVELS
 CC IN LIVER, LUNG AND KIDNEY.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- DISEASE: DEFECTS IN COL18A1 ARE A CAUSE OF KNOBLOCH SYNDROME (KS);
 CC AN AUTOSOMAL RECESSIVE DISORDER DEFINED BY THE OCCURRENCE OF HIGH
 CC MYOPIA, VITREORETINAL DEGENERATION WITH RETINAL DETACHMENT,
 CC MACULAR ABNORMALITIES AND OCCIPITAL ENCEPHALOCELE.
 CC -1- SIMILARITY: BELONGS TO THE MULTIPLEXIN FAMILY OF COLLAGENS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AF018081; AAC39658.1; -;
 DR EMBL: AF018082; AAC39659.1; -;
 DR EMBL: AL163302; CAB90482.1; -;
 DR EMBL: L22548; AAA51864.1; -;
 DR EMBL: AF184060; AAF03310.1; ALT_INIT.
 DR GLOSULEDB: P39060; -;
 DR MIM: 120328; -;
 DR MIM: 267750; -;
 DR InterPro: IPR000087; -;
 DR Pfam: PF01391; Collagen: 6.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing.
 FT SIGNAL
 FT CHAIN 1 23
 FT CHAIN 24 1516
 FT CHAIN 1334 1516
 FT CHAIN 24 516
 FT CHAIN 517 550
 FT CHAIN 551 560
 FT CHAIN 561 640
 FT CHAIN 641 664
 FT CHAIN 665 786
 FT CHAIN 787 809
 FT CHAIN 810 892
 FT CHAIN 810 892

FT DOMAIN 893 906
 FT DOMAIN 907 948
 FT DOMAIN 949 961
 FT DOMAIN 962 1034
 FT DOMAIN 1035 1044
 FT DOMAIN 1045 1077
 FT DOMAIN 1078 1089
 FT DOMAIN 1090 1111
 FT DOMAIN 1112 1118
 FT DOMAIN 1119 1173
 FT DOMAIN 1174 1186
 FT DOMAIN 1187 1204
 FT DOMAIN 1205 1516
 FT CARBOHYD 68 68
 FT CARBOHYD 129 129
 FT CARBOHYD 164 164
 FT CARBOHYD 691 691
 FT DISULFID 1366 1506
 FT DISULFID 1468 1498
 FT SITE 1095 1097
 FT VARSPIC 1 180
 FT VARSPIC 181 215
 FT CONFLICT 428 428
 FT CONFLICT 841 841
 FT CONFLICT 877 877
 FT CONFLICT 886 886
 FT CONFLICT 912 912
 FT CONFLICT 933 933
 FT CONFLICT 975 975
 FT CONFLICT 1064 1064
 FT CONFLICT 1084 1084
 FT CONFLICT 1120 1120
 FT CONFLICT 1123 1123
 FT CONFLICT 1126 1126
 FT CONFLICT 1206 1206
 FT CONFLICT 1304 1304
 FT CONFLICT 1314 1314
 FT CONFLICT 1323 1324
 FT CONFLICT 1437 1437
 FT CONFLICT 1443 1443
 FT CONFLICT 1483 1483
 FT SEQUENCE 1516 AA; 153840 MW; 3C70F29A4476BE76 CRC64;
 Query Match 92.7%; Score 38; DB 1; Length 1516;
 Best Local Similarity 87.5%; Pred. No. 6.7;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 OY 1 SYTYCTE 8
 DB 1501 AYTYCTE 1508
 RESULT 3
 YAI7_HUMAN STANDARD; PRT; 1015 AA.
 ID YAI7_HUMAN
 AC Q90P23;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN KIAA1017.
 GN KIAA1017.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eulheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Ref 6:63-70(1999).
CC -1- SIMILARITY: TO HUMAN KIAA0329.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AB023234; BAA76861.1; -
CC Hypothetical protein.
KW SEQUENCE 1015 AA; 114842 MW; 0214C2BFD67EA426 CRC64;
SO

OY 2 YVICIE 8
Db 875 YVICIE 881

Query Match 78.0%; Score 32; DB 1; Length 1015;
Best Local Similarity 57.1%; Pred. No. 70;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

RESULT 4
YA36_METJA STANDARD; PRT; 233 AA.
ID YA36_METJA
AC Q58442;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ1036.
GN MJ1036.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE-96337999; PubMed-8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
RA Scott J.L., Geoghegan N.S.M., Meldrum J.F., Fuhmann J.M., Glodek A.,
RA Ullrich-Back T.R., Kelley J.M., Weidman J.F., Sadow P.W., Hanna M.C.,
RA Klenk H.P., Roberts K.M., Hursel M.A., Kaine B.P., Borodovsky M.,
RA Clontz H., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U67546; AAB99045.1; -
CC TIGR: M1036; -
CC Hypothetical protein; Transmembrane.
KW
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 36 56 POTENTIAL.
FT TRANSMEM 62 82 POTENTIAL.
FT TRANSMEM 119 139 POTENTIAL.
SO

FT TRANSMEM 159 179 POTENTIAL.
FT TRANSMEM 188 208 POTENTIAL.
SO SEQUENCE 233 AA; 26170 MW; D2D82ACF620DB58 CRC64;

OY 1 SYVICI 7
Db 15 SYVICI 21

Query Match 73.2%; Score 30; DB 1; Length 233;
Best Local Similarity 71.4%; Pred. No. 47;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

RESULT 5
RDS_FELCA STANDARD; PRT; 346 AA.
ID RDS_FELCA
AC P35906;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PERIPHERIN (RETINAL DEGENERATION SLOW PROTEIN).
GN RDS.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ABYSSINIAN;
RX MEDLINE-94162777; PubMed-8118105;
RA Gorin M.B., Snyder S., To A.C., Narfstrom K., Curtis R.;
RT "The cat RDS transcript: candidate gene analysis and phylogenetic
RT sequence analysis."
RL Mamm. Genome 4:544-548(1993).
CC -1- FUNCTION: MAY FUNCTION AS AN ADHESION MOLECULE INVOLVED IN
CC STABILIZATION AND COMPACTION OF OUTER SEGMENT DISKS OR IN THE
CC MAINTENANCE OF THE CURVATURE OF THE RIM. IT IS ESSENTIAL FOR DISK
CC MORPHOGENESIS.
CC
CC -1- SUBUNIT: HOMODIMER, DISULFIDE LINKED. PROBABLY FORMS A COMPLEX
CC WITH A ROM1 HOMODIMER. OTHER PROTEINS COULD ASSOCIATE WITH THIS
CC COMPLEX IN RDS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: RETINA (PHOTORECEPTOR). IN RIM REGION OF RDS
CC (RDS OUTER SEGMENT) DISKS.
CC -1- SIMILARITY: BELONGS TO THE RDS(PERIPHERIN) / ROM1 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M94047; AAA19175.1; -
CC InterPro: IPR000830; -
DR Pfam: PF00824; RDS_ROM1.1.
DR PRINTS: PR00218; PERIPHERIN.RDS.
DR PROSITE: PS00930; RDS_ROM1.1.
KW Photoreceptor; Vision; Transmembrane; glycoprotein.
FT TRANSMEM 1 24 POTENTIAL.
FT TRANSMEM 25 43 POTENTIAL.
FT TRANSMEM 44 61 POTENTIAL.
FT TRANSMEM 62 80 POTENTIAL.
FT DOMAIN 81 99 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 100 123 POTENTIAL.
FT DOMAIN 124 264 LUMENAL (POTENTIAL).
FT TRANSMEM 265 290 LUMENAL (POTENTIAL).
FT DOMAIN 291 346 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 229 229 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 346 AA; 39171 MW; E70DF0E29E4759A6 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 346;
 Best Local Similarity 57.18; Pred. No. 67;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCI 7
 111111
 DB 100 SYIVLCV 106

RESULT 6
 RFC_SALTY STANDARD; PRT; 407 AA.
 ID P26479;
 AC 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE O-ANTIGEN POLYMERASE.
 GN RFC.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmonella.
 OX NCBI_Taxid=602;
 RN {1}
 RP SEQUENCE FROM N.A.
 RX MEDLINE-9193209; PubMed-1707412;
 RA Collins L.V., Hackett J.;
 RT "Molecular cloning, characterization, and nucleotide sequence of the
 rfc gene, which encodes an O-antigen polymerase of Salmonella
 typhimurium.";
 RL J. Bacteriol. 173:2521-2529(1991).
 CC -1- FUNCTION: MAY LINK THE O-ANTIGEN TETRASACCHARIDE UNITS INTO LONG
 CHAINS, GIVING RISE TO TYPICAL SMOOTH LPS.
 CC -1- PATHWAY: LIPIDPOLYSACCHARIDE BIOSYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
 CC CONTAINS SEVERAL POTENTIAL TRANSMEMBRANE DOMAINS (POTENTIAL).
 CC -1- SIMILARITY: TO OTHER O-ANTIGEN POLYMERASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M60066; AAA27210.1; -;
 DR PIR: A43672; A43672.
 DR StyGene: SG10352; rfc.
 KM Lipopolysaccharide biosynthesis; Transmembrane; Inner membrane.
 FT TRANSSEM 2 22 POTENTIAL.
 FT TRANSSEM 31 51 POTENTIAL.
 FT TRANSSEM 63 83 POTENTIAL.
 FT TRANSSEM 101 121 POTENTIAL.
 FT TRANSSEM 141 161 POTENTIAL.
 FT TRANSSEM 179 199 POTENTIAL.
 FT TRANSSEM 211 231 POTENTIAL.
 FT TRANSSEM 320 340 POTENTIAL.
 FT TRANSSEM 356 376 POTENTIAL.
 FT TRANSSEM 382 402 POTENTIAL.
 SQ SEQUENCE 407 AA; 47461 MW; AF187D6833D9CEDE CRC64;

Query Match 73.2%; Score 30; DB 1; Length 407;
 Best Local Similarity 71.46; Pred. No. 77;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYIVLCI 7
 111111
 DB 5 SYIVLCI 11

RESULT 7
 NAPI_HUMAN STANDARD; PRT; 420 AA.
 ID NAPI_HUMAN
 AC O96009;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE NAPSIN 1 PRECURSOR (EC 3.4.23.-) (NAPSIN A) (NAPA) (TA01/TA02)
 DE (ASPARTYL PROTEASE 4) (ASP 4) (ASP4).
 GN NAPI.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RN {1}
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung, and Kidney;
 RX MEDLINE-9902989; PubMed-9877162;
 RA Tatejell P.J., Powell D.J., Hill J., Smith T.S., Tew D.G., Kay J.;
 RT "Napsins: new human aspartic proteases. Distinction between two
 closely related genes.";
 RL FEBS Lett. 441:43-48(1998).
 RN {2}
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Koelsch G., Wu S., Henthorn J., Tang J., Lin X.;
 RT "New human aspartic proteases napsin 1 and napsin 2: molecular
 cloning and intracellular localization of napsin 1.";
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN {3}
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND MASS SPECTROMETRY.
 RC TISSUE=Fetal lung;
 RX MEDLINE-2004/840; PubMed-10580105;
 RA Chuman Y., Bergman A.-C., Ueno T., Saito S., Sakaguchi K.,
 RA Alaiya A.A., Franzen B., Bergman T., Arnott D., Auer G., Appella E.,
 RA Joernvall H., Linder S.;
 RT "Napsin A, a member of the aspartic protease family, is abundantly
 expressed in normal lung and kidney tissue and is expressed in lung
 adenocarcinomas.";
 RL FEBS Lett. 462:129-134(1999).
 RN {4}
 RP SEQUENCE FROM N.A.
 RA Blankowski M.J., Shuck M.E., Slightom J.L., Drong R.P.;
 RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY BE INVOLVED IN PROCESSING OF PNEUMOCYTE SURFACTANT
 PRECURSORS.
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN ADULT LUNG (TYPE II
 PNEUMOCYTES) AND KIDNEY AND IN FETAL LUNG. LOW LEVELS IN ADULT
 SPLEEN AND VERY LOW LEVELS IN PERIPHERAL BLOOD LEUKOCYTES.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
 BUKARTOTIC ASPARTYL PROTEASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF090386; AAD04917.1; -;
 DR EMBL: AF098484; AAD13215.1; -;
 DR EMBL: AF200345; AAF17081.1; -;
 DR MEROPS: A01.046; -;
 DR InterPro: IPR001461; -;
 DR InterPro: IPR001969; -;
 DR Pfam: PF00026; asp. 1.
 DR PRINTS: PR00792; PEPsin.
 DR PROSITE: PS00141; ASP_PROTEASE; 2.
 KM Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Signal.
 FT SIGNAL 1 25 POTENTIAL.
 FT PROPEP 26 63 ACTIVATION PEPTIDE.
 FT CHAIN 64 420 NAPSIN 1.

FT ACT_SITE 96 96 BY SIMILARITY.
 FT ACT_SITE 283 283 BY SIMILARITY.
 FT DISULFID 109 116 BY SIMILARITY.
 FT DISULFID 274 278 BY SIMILARITY.
 FT DISULFID 317 354 BY SIMILARITY.
 FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 420 AA; 45386 MW; 018B6AE5BD0C865 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 420;
 Best Local Similarity 71.4%; Pred. No. 79;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 2 YIVLCIE 8
 Db 313 YIILCE 319

RESULT 8
 ID NP73_HUMAN STANDARD: PRT: 436 AA.
 AC 000624;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE SODIUM-DEPENDENT PHOSPHATE TRANSPORT PROTEIN 3 (SODIUM/PHOSPHATE
 DE COTRANSPORTER 3) (NA(+)/PI COTRANSPORTER 3).
 GN SLC17A3 OR NP73.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ruddy D.A., Krommal G.S., Lee V.K., Mintler G.A., Quintana L.,
 RA Domingo R., Meyer N.C., Basava A., McClelland E., Fullan A.,
 RA Mapa F.A., Moore T., Thomas W., Loeb D.B., Harmon C., Tsuchihashi Z.,
 RA Wolff R.K., Schatman R.C., Feder J.N.,
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 CC -1 FUNCTION: MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE
 CC INTO CELLS VIA NA+ COTRANSPORT (BY SIMILARITY).
 CC -1 SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: U91328; AAB82085.1; -
 DR EMBL: U90544; AAB53422.1; -
 KM Transport; Transmembrane; Glycoprotein; Phosphorylation; Symport;
 KW Sodium transport.
 FT TRANSMEM 26 46 POTENTIAL.
 FT TRANSMEM 98 118 POTENTIAL.
 FT TRANSMEM 130 150 POTENTIAL.
 FT TRANSMEM 183 203 POTENTIAL.
 FT TRANSMEM 211 231 POTENTIAL.
 FT TRANSMEM 273 293 POTENTIAL.
 FT TRANSMEM 317 337 POTENTIAL.
 FT TRANSMEM 386 406 POTENTIAL.
 FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 69 69 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 436 AA; 47350 MW; DF02F618E83A572F CRC64;

Query Match 73.2%; Score 30; DB 1; Length 436;

Best Local Similarity 57.1%; Pred. No. 82;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SYIVLCI 7
 Db 193 SFYILCV 199

RESULT 9
 ID YMA40_YEAST STANDARD: PRT: 550 AA.
 AC 003212; Q03830;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL 62.5 KDA PROTEIN IN ALD2-DDP48 INTERGENIC REGION.
 GN YMR171C OR YMR610.01C OR YMR520.20C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE OF 1-241 FROM N.A.
 RA STRAIN-S288C / AB972;
 RA Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.,
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 215-550 FROM N.A.
 RA STRAIN-S288C / AB972;
 RA Hunt S., Bowman S., Barrell B.G., Rajandream M.A.,
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 CC -1 SIMILARITY: TO YEAST YKL124W.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: 249808; CAAB9904.1; -
 DR EMBL: 249705; CAAB9807.1; -
 DR SGD: S0004781; YMR171C.
 DR InterPro: IPR000107; -
 DR Pfam: PF00622; Spry; 1;
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 46 66 POTENTIAL.
 SO SEQUENCE 550 AA; 62532 MW; 6B1C6E9F9A889FA CRC64;

Query Match 73.2%; Score 30; DB 1; Length 550;
 Best Local Similarity 83.3%; Pred. No. 1e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 2 YIVLCI 7
 Db 58 YIVICI 63

RESULT 10
 ID COMP_BACSU STANDARD: PRT: 769 AA.
 AC 099027; Q05226;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 40, Last annotation update)
 DE SENSOR PROTEIN COMP (EC 2.7.1.-).
 GN COMP.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;

[1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BD630:
 RA MEDLINE-90337321: PubMed-2115363;
 RA Weinrauch Y., Penchev R., Dubnau D.;
 RT "A Bacillus subtilis regulatory gene product for genetic competence
 RT and sporulation resembles sensor protein members of the bacterial
 RT two-component signal-transduction systems.";
 RL Genes Dev. 4:860-872(1990).
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA Oudega B., Koningssteijn G., Van Zoest A.;
 RT "Bacillus subtilis genome project, DNA sequence from yufA to
 RT yufD.";
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE OF 1-16 FROM N.A.
 RC STRAIN-168;
 RA Weinrauch Y., Msadek T., Kunst F., Dubnau D.;
 RT "Sequence and properties of comQ, a new competence regulatory gene of
 RT Bacillus subtilis.";
 RL J. Bacteriol. 173:5685-5693(1991).
 CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM COMP/COMA.
 CC HISTIDINE KINASE THAT IS REQUIRED EARLY IN THE COMPETENCE CASCADE.
 CC ACTIVATES COMA PROTEIN BY PHOSPHORYLATION. IT PLAYS ROLE IN
 CC SPOREATION. AT LEAST PARTLY INTERCHANGEABLE WITH THAT OF SPOIIJ.
 CC -1- SIMILARITY: TO OTHER PROKARYOTIC REGULATORY PROTEINS WHICH BELONG
 CC TO A TWO-COMPONENT REGULATORY SYSTEM AND TRANSDUCE ENVIRONMENTAL
 CC SIGNALS TO TRANSCRIPTIONAL APPARATUS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X54010; CA37957.1; -
 DR EMBL: 293932; CAB07903.1; -
 DR EMBL: M22856; AAA22319.1; -
 DR EMBL: M71283; AAA22324.1; -
 DR EMBL: 299120; CAB15157.1; -
 DR PIR: A35848; A35848.
 DR SUBLIST: BG10380; comp.
 KW Sensory transduction; Phosphorylation; Transferase; Kinase;
 KW Transmembrane.
 FT DOMAIN 1 9
 FT TRANSMEM 10 33
 FT TRANSMEM 34 113
 FT TRANSMEM 114 134
 FT TRANSMEM 135 144
 FT TRANSMEM 145 167
 FT TRANSMEM 168 235
 FT TRANSMEM 236 257
 FT TRANSMEM 258 272
 FT TRANSMEM 273 295
 FT TRANSMEM 296 299
 FT TRANSMEM 300 323
 FT TRANSMEM 324 337
 FT TRANSMEM 338 357
 FT TRANSMEM 358 361
 FT TRANSMEM 362 383
 FT TRANSMEM 384 769
 FT DOMAIN 533 769
 FT MOD_RES 456 456
 FT CONFLICT 604 604
 FT CONFLICT 610 610
 FT CONFLICT 628 628
 FT CONFLICT 636 637
 FT CONFLICT 637 637
 OL -> PV (IN REF. 1).
 S -> C (IN REF. 1).
 D -> Y (IN REF. 1).
 E -> G (IN REF. 1).
 OL -> PV (IN REF. 1).

SEQ SEQUENCE 769 AA: 89316 MW: ESCEACE278DES5B4 CRC64;
 Query Match 73.2%; Score 30; DB 1; Length 769;
 Best Local Similarity 57.1%; Pred. No. 1,3e+02;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SYVLCT 7
 Db 497 SYVLCT 503
 RESULT 11
 TH11_SCHPO
 ID TH11_SCHPO STANDARD; PRT; 775 AA.
 AC P36598; P40378;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE THIAMINE REPRESSIBLE GENES REGULATORY PROTEIN TH11 (TRANSCRIPTION
 DE FACTOR NTF1).
 GN TH11 OR NTF1.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetaceae; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972;
 RA MEDLINE-94374699; PubMed-8088540;
 RA Frankhauser H., Schweingruber M.E.;
 RT "Thiamine-repressible genes in Schizosaccharomyces pombe are
 RT regulated by a Cys6 zinc-finger motif-containing protein.";
 RL Gene 147:141-144(1994).
 RP SEQUENCE OF 1-728 FROM N.A.
 RC STRAIN-972;
 RA MEDLINE-94216297; PubMed-8163491;
 RA Tang C.S.L., Bueno A., Russell P.;
 RT "ntf1 encodes a 6-cysteine zinc finger-containing transcription
 RT factor that regulates the nmt1 promoter in fission yeast.";
 RL J. Biol. Chem. 269:11921-11926(1994).
 CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT ACTIVATES THE NMT1 PROMOTER.
 CC REGULATION OF THIAMINE REPRESSIBLE GENES.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- SIMILARITY: CONTAINS A ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
 CC CLUSTER DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X77512; CA454648.1; -
 DR EMBL: L25912; AAA19010.1; -
 DR PIR: S41962; S41962.
 DR PIR: S43749; S43749.
 DR PIR: A53575; A53575.
 DR HSSP: P04386; 1D66.
 DR InterPro: IPR001138; -
 DR Pfam: PF001172; ZN_Clus; 1.
 DR PROSITE: PS00463; ZN2_CYS6_FUNGAL_1; 1.
 DR PROSITE: PS50048; ZN2_CYS6_FUNGAL_2; 1.
 KW Transcription regulation; Activator; DNA-binding; Nuclear protein;
 KW Zinc; Metal-binding.
 FT DNA_BIND 39 65
 FT CONFLICT 10 10
 FT CONFLICT 325 326
 FT CONFLICT 684 684
 FT CONFLICT 684 684
 S -> C (IN REF. 2).
 N -> K (IN REF. 2).
 SA -> LR (IN REF. 2).
 S -> C (IN REF. 2).

SQ SEQUENCE 775 AA: 88060 MW: 88E2A48072F678FB CRC64:
 Query Match 73.2%; Score 30; DB 1; Length 775;
 Best Local Similarity 71.4%; Pred. No. 1.4e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 YIVLCI 8
 DB 547 YIVLCVE 553

RESULT 12
 VP_H1_NEUCR STANDARD: PRT: 856 AA.
 ID VP_H1_NEUCR 001290:
 AC 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE VACUOLAR ATP SYNTHASE 98 KDA SUBUNIT (EC 3.6.1.34) (VACUOLAR ATPASE 98 KDA SUBUNIT).
 GN VP_H1.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_Taxid=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=74-OR23-1A;
 RA Bowman E.J.;
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR ASSEMBLY AND ACTIVITY OF THE VACUOLAR
 ATPASE. POTENTIAL ROLE IN DIFFERENTIAL TARGETING AND REGULATION OF
 THE ENZYME FOR A SPECIFIC ORGANELLE (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. VACUOLAR.
 CC -1- SIMILARITY: BELONGS TO THE VAPASE 116 KDA SUBUNIT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 or send an email to license@sdb-sib.ch).
 CC -----
 DR EMBL: U36396; AAA93078.1; -
 DR InterPro: IPR002430; -
 DR Pfam: PF01496; V_ATPase_sub.a; 1.
 KW Hydrogen ion transport; Transmembrane; Glycoprotein.
 FT DOMAIN 1 415
 FT TRANSMEM 416 436
 FT DOMAIN 437 457
 FT TRANSMEM 458 478
 FT DOMAIN 479 548
 FT TRANSMEM 549 569
 FT DOMAIN 570 581
 FT TRANSMEM 582 602
 FT DOMAIN 603 615
 FT TRANSMEM 616 636
 FT DOMAIN 637 642
 FT TRANSMEM 643 663
 FT DOMAIN 664 754
 FT TRANSMEM 755 775
 FT DOMAIN 776 786
 FT TRANSMEM 787 807
 FT DOMAIN 808 856
 FT CAROXYD 119
 FT SEQUENCE 856 AA: 97992 MW: F75E3737ADAD3D8 CRC64:

Query Match 73.2%; Score 30; DB 1; Length 856;
 Best Local Similarity 83.3%; Pred. No. 1.5e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 YIVLCI 7
 DB 595 YIVLCI 600

RESULT 13
 CAIE_HUMAN STANDARD: PRT: 1388 AA.
 ID CAIE_HUMAN P39059:
 AC 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE COLLAGEN ALPHA 1(XV) CHAIN PRECURSOR.
 GN COL15A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=umbilical cord;
 RX MEDLINE=94148920; PubMed=8106446;
 RA Kivirikko S., Heineken P., Renk M.V., Honkanen N., Myers J.C.,
 RA Pihlajaniemi T.;
 RT "Primary structure of the alpha 1 chain of human type XV collagen and
 exon-intron organization in the 3' region of the corresponding
 gene."
 RT J. Biol. Chem. 269:4773-4779(1994).
 RN [2]
 RP SEQUENCE OF 1-569 FROM N.A.
 RC TISSUE=placenta;
 RX MEDLINE=94140817; PubMed=8307960;
 RA Murgaki Y., Abe N., Nishimura Y., Olsen B.R., Ooshima A.;
 RT "The human alpha 1(XV) collagen chain contains a large amino-terminal
 non-triple helical domain with a tandem repeat structure and homology
 to alpha 1(XVII) collagen."
 RT J. Biol. Chem. 269:4042-4046(1994).
 RN [3]
 RP SEQUENCE OF 544-1252 FROM N.A.
 RX MEDLINE=93066196; PubMed=1279671;
 RA Myers J.C., Kivirikko S., Gordon M.K., Dion A.S., Pihlajaniemi T.;
 RT "Identification of a previously unknown human collagen chain, alpha
 1(XV), characterized by extensive interruptions in the triple-helical
 region."
 RT Proc. Natl. Acad. Sci. U.S.A. 89:10144-10148(1992).
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN INTERNAL ORGANS
 SUCH AS ADRENAL GLAND, PANCREAS AND KIDNEY.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- SIMILARITY: BELONGS TO THE MULTIPLEXIN FAMILY OF COLLAGENS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 or send an email to license@sdb-sib.ch).
 CC -----
 DR EMBL: U2586; AAA58429.1; -
 DR EMBL: D21230; BAA04762.1; -
 DR EMBL: L01697; -; NOT_ANNOTATED_CDS.
 DR MIM: 120325; -
 DR InterPro: IPR000087; -
 DR Pfam: PF01391; Collagen; 4.
 KW Cell adhesion; Collagen; Glycoprotein; Signal.
 FT SIGNAL 1 25
 FT CHAIN 26 1388
 FT DOMAIN 26 555
 FT DOMAIN 556 573
 FT DOMAIN 574 618
 FT DOMAIN 619 732
 FT TRIPLE-HELICAL REGION 1 (NC1).
 FT TRIPLE-HELICAL REGION 1 (COL1).
 FT TRIPLE-HELICAL REGION 2 (NC2).
 FT TRIPLE-HELICAL REGION 2 (COL2).


```

FT DOMAIN 733 763 NONHELICAL REGION 3 (NC3).
FT DOMAIN 764 798 TRIPLE-HELICAL REGION 3 (COL3).
FT DOMAIN 799 822 TRIPLE-HELICAL REGION 4 (NC4).
FT DOMAIN 823 867 TRIPLE-HELICAL REGION 4 (COL4).
FT DOMAIN 868 878 TRIPLE-HELICAL REGION 5 (NC5).
FT DOMAIN 879 949 TRIPLE-HELICAL REGION 5 (COL5).
FT DOMAIN 950 983 TRIPLE-HELICAL REGION 6 (NC6).
FT DOMAIN 984 1013 TRIPLE-HELICAL REGION 6 (COL6).
FT DOMAIN 1014 1027 TRIPLE-HELICAL REGION 7 (NC7).
FT DOMAIN 1028 1045 TRIPLE-HELICAL REGION 7 (COL7).
FT DOMAIN 1046 1052 TRIPLE-HELICAL REGION 8 (NC8).
FT DOMAIN 1053 1107 TRIPLE-HELICAL REGION 8 (COL8).
FT DOMAIN 1108 1117 TRIPLE-HELICAL REGION 9 (NC9).
FT DOMAIN 1118 1132 TRIPLE-HELICAL REGION 9 (COL9).
FT DOMAIN 1133 1388 TRIPLE-HELICAL REGION 10 (NC10).
FT DOMAIN 358 408 4 X TANDEM REPEATS.
FT REPEAT 409 459 1.
FT REPEAT 460 509 2.
FT REPEAT 510 555 3.
FT CARBOHYD 306 306 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 687 687 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 814 814 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1046 1046 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 10 10 C -> S (IN REF. 2).
FT CONFLICT 49 49 D -> V (IN REF. 2).
FT CONFLICT 95 95 L -> A (IN REF. 2).
FT CONFLICT 150 150 P -> A (IN REF. 2).
FT CONFLICT 204 204 M -> V (IN REF. 2).
FT CONFLICT 409 409 R -> A (IN REF. 2).
SQ SEQUENCE 1388 AA; 141930 MW; 60822AD925A3093D CRC64;

```

```

Query Match 73.2%; Score 30; DB 1; Length 1388;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 3 YIVCIE 8
Db 1374 YIVCIE 1379

```

```

RESULT 14
YOTO CAEEL
ID YOTO CAEEL STANDARD; PRT; 80 AA.
AC P34655;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOPHETICAL 8.7 KDA PROTEIN ZK632.10 IN CHROMOSOME III.
GN ZK632.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodermidae; Caenorhabditis.
OX NCBI_TaxID=6233;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton L., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Ritken L., Roopka A., Saunders D., Showkhen M.,
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Watson R., Watson A., Welnstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
*2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

```

```

RT elegans.*
RL Nature 368:32-38(1994).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE UPF0057 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z22181; CA80190.1; -.
DR PIR; S40942; S40942.
DR Wormpep; ZK632.10; CE02385.
DR InterPro; IPR000612; -.
DR Pfam; PF01679; UPF0057.1.
DR PROSITE; PS01309; UPF0057.1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 4 24 POTENTIAL.
FT TRANSMEM 32 52 POTENTIAL.
SQ SEQUENCE 80 AA; 8708 MW; 3741ABE9B2DALA33 CRC64;

```

```

Query Match 70.7%; Score 29; DB 1; Length 80;
Best Local Similarity 71.4%; Pred. No. 29;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 2 YIVCIE 8
Db 48 YIVCIE 54

```

```

RESULT 15
MCSR HUMAN
ID MCSR HUMAN STANDARD; PRT; 325 AA.
AC P33032;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MELANOCORTIN-5 RECEPTOR (MCS-R) (MC-2).
GN MCSR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Placenta;
RX MEDLINE=93384614; PubMed=8396929;
RA Chhajlani V., Muceniece R., Wikberg J.E.S.;
RT "Molecular cloning of a novel human melanocortin receptor.*";
RL Biochem. Biophys. Res. Commun. 195:866-873(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94234987; PubMed=8179577;
RA Griffon N., Mignon V., Facchinetti P., Diaz J., Schwartz J.C.,
RA Sokoloff P.;
RT "Molecular cloning and characterization of the rat fifth melanocortin
RT receptor.*";
RL Biochem. Biophys. Res. Commun. 200:1007-1014(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-Placenta;
RX MEDLINE=95258173; PubMed=7739752;
RA Fath Z., Iben L.G., Parker E.M.;
RT "Cloning, expression, and tissue distribution of a fifth melanocortin
RT receptor subtype.*";
RL Neurochem. Res. 20:107-113(1995).
CC -1- FUNCTION: RECEPTOR FOR MSH (ALPHA, BETA AND GAMMA) AND ACTH. THE
CC ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
CC ADENYLATE CYCLASE. THIS RECEPTOR IS A POSSIBLE MEDIATOR OF THE

```

```
CC      IMMUNOMODULATION PROPERTIES OF MELANOCORTINS.
CC      -1- SUBCELLULAR LOCATION: INTERAL MEMBRANE PROTEIN.
CC      -1- TISSUE SPECIFICITY: IS EXPRESSED IN THE BRAIN BUT NOT IN THE
CC      MELANOMA CELLS.
CC      -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; 225470; CAA80962.1; -
DR      EMBL; 127080; AAA59566.1; -
DR      EMBL; U08353; AAB60376.1; -
DR      GCRDB; GCR_0793; -
DR      GCRDB; GCR_0840; -
DR      GCRDB; GCR_2058; -
DR      MIM; 600042; -
DR      InterPro; IPR000276; -
DR      InterPro; IPR000621; -
DR      InterPro; IPR001671; -
DR      InterPro; IPR001908; -
DR      Pfam; PF00001; 7tm_1; 2.
DR      PRINTS; PR00534; MCRFAMILY.
DR      PRINTS; PR00535; MELNOCORTINR.
DR      PRINTS; PR01063; MELNOCORTINR.
DR      PROSITE; PS00237; G_PROTEIN_RECEP_FL_1; 1.
DR      PROSITE; PS50262; G_PROTEIN_RECEP_FL_2; 1.
KW      G-protein coupled receptor; Transmembrane; Glycoprotein;
KW      Phosphorylation; Lipoprotein; Palmitate.
FT      DOMAIN 1 37
FT      TRANSMEM 38 61
FT      DOMAIN 62 73
FT      TRANSMEM 74 97
FT      DOMAIN 98 114
FT      TRANSMEM 115 138
FT      DOMAIN 139 155
FT      TRANSMEM 156 179
FT      DOMAIN 180 186
FT      TRANSMEM 187 211
FT      DOMAIN 212 239
FT      TRANSMEM 240 265
FT      DOMAIN 266 273
FT      TRANSMEM 274 297
FT      DOMAIN 298 325
FT      CARBOHYD 2 2
FT      CARBOHYD 15 15
FT      CARBOHYD 20 20
FT      CARBOHYD 28 28
FT      LIPID 311 311
FT      LIPID 312 312
FT      CONFLICT 149 149
FT      CONFLICT 221 234
FT      CONFLICT 297 297
SQ      SEQUENCE 325 AA; 36600 MW; 8BEC17E1BDA059BB CRC64;
      F -> Y (IN REF. 2).
      ALPGASSARQRTSM -> LCPGPAALRGSPAW (IN
      REF. 1).
```

Query Match 70.7%: Score 29; DB 1; Length 325;

Best Local Similarity 42.9%: Pred. No. 99;

Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCI 7

DB 184 TYVILCL 190

Search completed: August 16, 2001, 11:44:17
Job time: 169 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 11:38:57 ; Search time 12.81 seconds
(without alignments)
47.572 Million cell updates/sec

Title: US-09-589-777A-2_COPY_168_175

Perfect score: 41

Sequence: 1 SYIVLCIE 8

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR.68:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	1315	2 A56101	collagen alpha 1(X
2	41	100.0	1774	2 B56101	collagen alpha 1(X
3	38	92.7	684	2 A53019	collagen alpha 1(X
4	32	78.0	147	2 F71498	hypothetical prote
5	32	78.0	147	2 E81656	conserved hypotet
6	32	78.0	293	2 D96505	unknown protein fl
7	32	78.0	1142	2 A45031	cysteine-rich fibr
8	31	75.6	1175	2 S52417	E-selectin ligand-
9	31	75.6	39	2 G85837	hypothetical prote
10	31	75.6	158	2 H72029	conserved hypotet
11	31	75.6	158	2 D86593	hypothetical prote
12	31	75.6	158	2 D84853	hypothetical prote
13	31	75.6	235	2 T23501	hypothetical prote
14	31	75.6	316	2 B75323	conserved hypotet
15	31	75.6	466	2 B82482	alpha-amylase VCAO
16	31	75.6	479	2 B86285	hypothetical prote
17	31	75.6	566	2 E81404	acetylactate synth
18	31	75.6	2342	2 T18200	faty-acid synthas
19	30	73.2	96	2 T42273	hypothetical prote
20	30	73.2	129	2 T21687	hypothetical prote
21	30	73.2	141	2 C71605	hypothetical prote
22	30	73.2	151	2 T48823	clathrin coat asse
23	30	73.2	163	2 T33130	hypothetical prote
24	30	73.2	168	2 T12834	hypothetical prote
25	30	73.2	233	2 C64429	hypothetical prote
26	30	73.2	233	2 D83313	probable permease
27	30	73.2	277	2 D84596	hypothetical prote
28	30	73.2	346	2 I46087	peripherin - cat
29	30	73.2	353	2 G71567	probable cytochrom

ALIGNMENTS

RESULT 1
A56101
collagen alpha 1(XVIII) chain precursor, short splice form - mouse
N:Contains: endostatin
C:Species: Mus musculus (house mouse)
C>Date: 03-Oct-1995 #sequence-revision 08-May-1998 #text-change 31-Mar-2000
C/Accession: A56101; A58371; S72450; S65595
J:Rehm, M.; Philajantemi, T.
J. Biol. Chem. 270, 4705-4711, 1995
A>Title: Identification of three N-terminal ends of type XVIII collagen chains and t
lif homologous to rat and Drosophila fibrillated proteins.
A:Reference number: A56101; M01D:95181468
A:Accession: A56101
A:Molecule type: mRNA
A:Residues: 1-103 <REH1>
A:Cross-references: GB:U11636; NID:9618427; PIDN:AA052178.1; PID:9618428
R:Rehm, M.; Philajantemi, T.
Proc. Natl. Acad. Sci. U.S.A. 91, 4224-4238, 1994
A>Title: Alpha1(XVIII), a collagen chain with frequent interruptions in the collageno
A:Reference number: A58371; M01D:94240112
A:Accession: A58371
A:Molecule type: mRNA
A:Residues: 1-928 <REH2>
A:Cross-references: GB:U16898; NID:9404754; PIDN:AAA7434.1; PID:9553894
R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
Submitted to the EMBL Data Library, August 1993
A:Reference number: S72450
A:Accession: S72450
A:Molecule type: mRNA
A:Residues: 28-687, 'L', 689-734, 'F', 736-751, 'R', 753-1315 <OHM>
A:Cross-references: EMBL:L22545; NID:9348968; PIDN:AAA19787.1; PID:9511298
R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
A>Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-
A:Reference number: A58370; M01D:94240111
A:Accession: S65595
A:Molecule type: mRNA
A:Residues: 28-1315 <OHS>
A:Cross-references: EMBL:L22545
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
lated and subsequently O-glycosylated.
C:Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in p
C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of
ay be useful in treating solid tumors.
C:Genetics:
A:Gene: MGI:Col18a1
A:Cross-references: MGI:71175
A:Map position: 10:41.0
C:Superfamily: unassigned collagens
C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteog
F1-25/Domain: signal sequence #status predicted <SIG>
F1-24-235/Region: thrombospondin amino-terminal similarity

```

F:26-1315/Product: collagen alpha 1(XVIII) chain, short splice form #status predicted <C>
F:327-353/Domains: collagenous #status predicted <C01>
F:364-433/Domains: collagenous #status predicted <C02>
F:462-583/Domains: collagenous #status predicted <C03>
F:607-688/Domains: collagenous #status predicted <C04>
F:704-743/Domains: collagenous #status predicted <C05>
F:759-831/Domains: collagenous #status predicted <C06>
F:842-874/Domains: collagenous #status predicted <C07>
F:887-910/Domains: collagenous #status predicted <C08>
F:892-894/Region: cell attachment (R-G-D) motif
F:918-969/Domains: collagenous #status predicted <C09>
F:983-1000/Domains: collagenous #status predicted <C010>
F:1132-1315/Product: endostatin #status predicted <EST>
F:1139-1315/Region: multiplexin collagen carboxyl-terminal similarity
F:126-488/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:172-228/Dissulfide bonds: #status predicted
F:240-245,1257/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:451,454,594/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match      100.0%; Score 41; DB 2; Length 1315;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SYVLICIE 8
Db 1299 SYVLICIE 1306

RESULT 2
B56101
collagen alpha 1(XVIII) chain precursor, long splice form - mouse
M:Contains: collagen alpha 1(XVIII) chain precursor, medium splice form; endostatin
C:Species: Mus musculus (house mouse)
C:Date: 03-Oct-1995 #sequence revision 08-May-1998 #text change 31-Mar-2000
C:Accession: B56101; C56101; S72450; S65595; P0675; A54072; A58816
R:Rehn, M.; Philantoniemi, T.
J. Biol. Chem. 270, 4705-4711, 1995
A:Title: Identification of three N-terminal ends of type XVIII collagen chains and their
clustering homologous to rat and Drosophila fibrinogen proteins.
A:Reference number: A56101; MUID:95181468
A:Accession: B56101
A:Molecule type: mRNA
A:Residues: 1-562 <REH>
A:Cross-references: GB:U11637; NID:9618429; PIDN:AA52179.1; PID:9618430
A:Experimental source: splice form clone PEL17.24
A:Accession: C56101
A:Molecule type: mRNA
A:Residues: 1-239,487-562 <REH2>
A:Cross-references: GB:U11637; NID:9618429
A:Experimental source: splice form clones PEL1, PEL9, PEL15, 2
R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
submitted to the EMBL Data Library, August 1993
A:Reference number: S72450
A:Accession: S72450
A:Molecule type: mRNA
A:Residues: 487-1146, 'L', 1148-1193, 'F', 1195-1210, 'R', 1212-1512, 'L', 1514-1522, 'F', 1524-16
A:Cross-references: EMBL:L22545; NID:9348968; PIDN:AA19787.1; PID:9511298
R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
A:Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa
A:Reference number: A58370; MUID:94240111
A:Accession: S65595
A:Molecule type: mRNA
A:Residues: 487-1512, 'L', 1514-1522, 'F', 1524-1683, 'V', 1685-1774 <OH2>
A:Cross-references: EMBL:L22545
R:Abbe, N.; Muragaki, Y.; Yoshioke, H.; Inoue, H.; Niinomiya, Y.
Biochem. Biophys. Res. Commun. 196, 576-582, 1993
A:Title: Identification of a novel collagen chain represented by extensive interruptions
A:Reference number: P0675; MUID:94059075
A:Accession: P0675
A:Molecule type: mRNA
A:Residues: 635-1774 <ABE>

```

```

R:Rehn, M.; Hantikka, E.; Philantoniemi, T.
J. Biol. Chem. 269, 13929-13935, 1994
A:Title: Primary structure of the alpha chain of mouse type XVIII collagen, partial
collagen chain.
A:Reference number: A54072; MUID:94245707
A:Accession: A54072
A:Molecule type: DNA; mRNA
A:Residues: 1293-1403, 'R', 1405-1774 <REH3>
A:Cross-references: GB:U03714; NID:9487733; PIDN:AA20657.1; PID:9487734
R:O'Reilly, M.S.; Boehm, T.; Shing, Y.; Fukai, N.; Vasios, G.; Lane, W.S.; Flynn, E.;
Cell 88, 277-285, 1997
A:Title: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.
A:Reference number: A58816; MUID:97160848
A:Accession: A58816
A:Molecule type: protein
A:Residues: 1591-1610 <ORE>
A:Experimental source: hemangioendothelium cells
A:Note: Inhibits endothelial cell proliferation
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
lated and subsequently O-glycosylated.
C:Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in p
C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of
ay be useful in treating solid tumors.
C:Genetics:
A:Gene: MGI:Coll8a1
A:Cross-references: MGI:71175
A:Map position: 10:41.0
A:Introns: 1295/3; 1310/1; 1331/1; 1345/3; 1388/3; 1437/1; 1461/3; 1505/3; 1516/3; 15
A:Note: the list of introns is incomplete
C:Superfamily: unassigned collagens
C:Keywords: alternative splicing; collagen alpha 1(XVIII) chain precursor, long splice form
F:1-239,487-1774/Product: collagen alpha 1(XVIII) chain precursor, medium splice form
F:1-24/Domains: signal sequence #status predicted <SIG>
F:361-488/Region: fibrinogen similarity
F:786-812/Domains: collagenous #status predicted <C001>
F:823-896/Domains: collagenous #status predicted <C002>
F:921-1042/Domains: collagenous #status predicted <C003>
F:1066-1148/Domains: collagenous #status predicted <C004>
F:1163-1204/Domains: collagenous #status predicted <C005>
F:1218-1290/Domains: collagenous #status predicted <C006>
F:1301-1333/Domains: collagenous #status predicted <C007>
F:1346-1369/Domains: collagenous #status predicted <C008>
F:1351-1353/Region: cell attachment (R-G-D) motif
F:1377-1428/Domains: collagenous #status predicted <C009>
F:1442-1459/Domains: collagenous #status predicted <C010>
F:1591-1774/Product: endostatin #status predicted <EST>
F:1598-1774/Region: multiplexin collagen carboxyl-terminal similarity
F:354,361,947/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:699,704,1716/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:910,913,1053/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match      100.0%; Score 41; DB 2; Length 1774;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SYVLICIE 8
Db 1758 SYVLICIE 1765

RESULT 3
A53019
collagen alpha 1(XVIII) chain - human (fragment)
M:Contains: endostatin
C:Species: Homo sapiens (man)
C:Date: 18-Aug-1995 #sequence revision 18-Aug-1995 #text change 31-Mar-2000
C:Accession: A53019
R:Oh, S.P.; Watanabe, M.L.; Seidlin, M.F.; Cheng, S.D.; Knoll, J.H.M.; Timmons, S.; Olse
Genomics 19, 494-499, 1994
A:Title: Cloning of cDNA and genomic DNA encoding human type XVIII collagen and local
A:Reference number: A53019; MUID:94245237

```

A:Accession: A53019
 A:Molecule type: mRNA
 A:Residues: 1-684 <ORF>
 A:Cross-references: GB:122548; NID:9348908; PIDN:AA51864.1; PID:9562794
 A:Note: The cited accession number, 123548, is not in Genbank release 103
 A:Note: In the authors' translation, 482-614 is not shown, residues 483-490 are shifted
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
 C:Comment: Different splice forms of collagen alpha 1(XVIII) may be involved in perivas
 C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un
 C:Comment: may be useful in treating solid tumors.
 C:Genetics:
 A:Gene: GDB:COL18A1
 A:Cross-references: GDB:138752; OMIM:120328
 A:Map position: 21q22.3-21q22.3
 C:Superfamily: unassigned collagens
 C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc
 F:1-684/Product: collagen alpha 1(XVIII) chain (fragment) #status predicted <MAT>
 F:1-59/Domin: collagenous (fragment) #status predicted <CO5>
 F:74-115/Domin: collagenous #status predicted <CO4>
 F:129-201/Domin: collagenous #status predicted <CO6>
 F:212-244/Domin: collagenous #status predicted <CO7>
 F:257-278/Domin: collagenous #status predicted <CO8>
 F:263-264/Region: collagenous #status predicted <CO9>
 F:286-340/Region: cell attachment (R-G-D) motif
 F:354-371/Domin: collagenous #status predicted <CO10>
 F:502-684/Product: endostatin #status predicted <EST>
 F:509-684/Region: multiplexin collagen carboxyl-terminal similarity

Query Match 92.7%; Score 38; DB 2; Length 684;
 Best Local Similarity 87.5%; Pred. No. 6.5;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCI 8
 Db 669 AYIVLCIE 676

RESULT 4
 F71498
 hypothetical protein C7565 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
 C:Species: Chlamydia trachomatis
 C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
 C:Accession: F71498
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
 Science 282, 754-759, 1998
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac
 A:Reference number: A71570; MUID:99000809
 A:Accession: F71498
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-147 <ARN>
 A:Cross-references: GB:AE001327; GB:AE001273; NID:93328999; PIDN:AAC68167.1; PID:9332900
 A:Experimental source: serotype D, strain UW-3/Cx
 C:Genetics:
 A:Gene: C7565

Query Match 78.0%; Score 32; DB 2; Length 147;
 Best Local Similarity 85.7%; Pred. No. 27;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCI 7
 Db 123 SFTVLCI 129

RESULT 5
 E81656
 conserved hypothetical protein TC0854 [imported] - Chlamydia muridarum (strain Nigg)
 C:Species: Chlamydia muridarum, Chlamydia trachomatis Mohn
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000

C:Accession: E81656
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
 C:Species: Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salze
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis Mohn and Chlamydia pneumoniae AR39
 A:Reference number: A81500; MUID:20150255
 A:Accession: E81656
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-147 <TEF>
 A:Cross-references: GB:AE002352; GB:AE002160; NID:97190879; PIDN:AAF39650.1; PID:9719
 A:Experimental source: strain Nigg (MOPn)
 C:Genetics:
 A:Gene: TC0854

Query Match 78.0%; Score 32; DB 2; Length 147;
 Best Local Similarity 85.7%; Pred. No. 27;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCI 7
 Db 123 SFTVLCI 129

RESULT 6
 D96505
 unknown protein F13N6.19 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: D96505
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Malt, R.; Marzla
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: D96505
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-293 <STO>
 A:Cross-references: GB:AE005173; NID:911024843; PIDN:AAG26928.1; GSPDB:GNC0141
 C:Genetics:
 A:Gene: F13N6.19
 A:Map position: 1

Query Match 78.0%; Score 32; DB 2; Length 293;
 Best Local Similarity 85.7%; Pred. No. 49;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYIVLCI 7
 Db 122 SYIALCI 128

RESULT 7
 A45031
 cysteine-rich fibroblast growth factor receptor - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
 C:Accession: A45031
 R:Burrus, L.W.; Zuber, M.E.; Lueddecke, B.A.; Olwin, B.B.
 Mol. Cell Biol. 12, 5600-5609, 1992
 A:Title: Identification of a cysteine-rich receptor for fibroblast growth factors.
 A:Reference number: A45031; MUID:93078761
 A:Accession: A45031
 A:Status: preliminary

A:Molecule type: mRNA; protein
A:Residues: 1-1142 <B0R>
A:Cross-references: GB:M95766; NID:g211775; PIDN:AAA48769.1; PID:g211776
A:Experimental source: embryos, brain
A>Note: sequence extracted from NCBI backbone (NCBIN:119082, NCBI:P:119083)
C:Superfamily: Caenorhabditis elegans hypothetical protein F14E5.2
C:Keywords: growth factor receptor

Query Match 78.0%; Score 32; DB 2; Length 1142;
Best Local Similarity 50.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCIE 8
||:|:|:|
Db 362 SYIVLCIE 369

RESULT 8
S52417
E-selectin ligand-1 protein - mouse
C:Species: Mus musculus (house mouse)
C>Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-2000
C:Accession: S52417
R:Steeegmaler, M.; Levinovitz, A.; Isenmann, S.; Borges, E.; Lenter, M.; Kocher, H.P.; Kl
Nature 373, 615-620, 1995
A:Title: The E-selectin-ligand ESL-1 is a variant of a receptor for fibroblast growth fa
A:Reference number: S52417; MUID:95157635
A:Accession: S52417
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1175 <STE>
A:Cross-references: EMBL:X84037; NID:g673435; PIDN:CA58855.1; PID:g673436
C:Superfamily: Caenorhabditis elegans hypothetical protein F14E5.2

Query Match 78.0%; Score 32; DB 2; Length 1175;
Best Local Similarity 50.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCIE 8
||:|:|:|
Db 395 SYIVLCIE 402

RESULT 9
G85837
hypothetical protein 23270 [Imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: G85837
R:Perina, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; DiMantana, E.; Potamoustis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: G85837
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-39 <STO>
A:Cross-references: GB:AE00174; NID:g12516312; PIDN:AA657163.1; GSPDB:GN00145; UMG:P:232
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: 23270

Query Match 75.6%; Score 31; DB 2; Length 39;
Best Local Similarity 57.1%; Pred. No. 13;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 YIVLCIE 8
::|:|:|
Db 19 FVIVLCIE 25

RESULT 10
H72029
conserved hypothetical protein CP1049 [Imported] - Chlamydia pneumoniae (strains

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: H72029; A81508
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: H72029

A:Molecule type: DNA
A:Residues: 1-158 <ARN>
A:Cross-references: GB:AE001663; GB:AE001363; NID:g4377134; PIDN:AA018971.1; PID:g437
A:Experimental source: strain CWL029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gail, S.R.; Heidelberg, J.F.; White, O.; Hicke
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MOPn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255
A:Accession: A81508
A:Molecule type: DNA
A:Residues: 1-158 <REA>
A:Cross-references: GB:AE002262; GB:AE002161; NID:g7189961; PIDN:AAE38822.1; PID:g718
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0822; CP1049

Query Match 75.6%; Score 31; DB 2; Length 158;
Best Local Similarity 71.4%; Pred. No. 45;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCI 7
|:|:|:|
Db 122 SFIVLCV 128

RESULT 11
D86593
CT965 hypothetical protein [Imported] - Chlamydia pneumoniae (strain J138)
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: D86593
R:Shital, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349
A:Accession: D86593
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-158 <STO>
A:Cross-references: GB:BA000008; NID:g9879196; PIDN:BA09030.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: CPJ0822

Query Match 75.6%; Score 31; DB 2; Length 158;
Best Local Similarity 71.4%; Pred. No. 45;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCI 7
|:|:|:|
Db 122 SFIVLCV 128

RESULT 12
D84853
hypothetical protein At2g42390 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: D84853
 R:Lin. X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffitt, K.S.; Cronin, L.A.; Shen, M.; Vankken, S.E.; Unayam, L.; Tallon, L.;
 Euse, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: AB4420; MUID:20083487
 A:Accession: D84853
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-163 <STO>
 A:Cross-references: GB:AE002093; NID:94567311; PIDN:AMD23722.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g42390
 A:Map position: 2

Query Match 75.6%; Score 31; DB 2; Length 163;
 Best Local Similarity 62.5%; Pred. No. 46;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYIVLCIE 8
 ||:||||
 DB 7 SYIVLCIE 14

RESULT 13
 T23501
 hypothetical protein K08G2.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Jun-2000
 C:Accession: T23501
 R:Lloyd, C.
 submitted to the EMBL Data Library, October 1996
 A:Reference number: 219749
 A:Accession: T23501
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-235 <WIL>
 A:Cross-references: EMBL:281100; PIDN:CAB03191.1; GSPDB:GN00023; CESP:K08G2.2
 C:Genetics:
 A:Experimental source: clone K08G2
 A:Gene: CESP:K08G2.2
 A:Map position: 5
 A:Introns: 52/2; 92/1; 110/3
 C:Superfamily: Caenorhabditis hypothetical protein C49G7.2

Query Match 75.6%; Score 31; DB 2; Length 235;
 Best Local Similarity 71.4%; Pred. No. 64;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLCI 7
 :|||||
 DB 133 NYIVICI 139

RESULT 14
 B75323
 conserved hypothetical protein - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C:Accession: B75323
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Uitterback, T.; Zalewski, C.; Mc
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: AV5250; MUID:20036896
 A:Accession: B75323
 A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-316 <WHI>
 A:Cross-references: GB:AE002040; GB:AE000513; NID:96459824; PIDN:AAF11588.1; PID:9645
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR2041
 A:Map position: 1

Query Match 75.6%; Score 31; DB 2; Length 316;
 Best Local Similarity 62.5%; Pred. No. 82;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SYIVLCIE 8
 ||:||||
 DB 113 SYIVLCIE 120

RESULT 15
 B82482
 alpha-amylase VCA0250 [Imported] - Vibrio cholerae (strain N16961 serogroup O1)
 C:Species: Vibrio cholerae
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C:Accession: B82482
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
 chardson, D.; Ermolaeva, M.D.; Venter, J.C.; Fraser, C.M.
 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A:Reference number: A82035; MUID:20406833
 A:Accession: B82482
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-466 <HEI>
 A:Cross-references: GB:AE004365; GB:AE003853; NID:9657642; PIDN:AAF96161.1; GSPDB:GN
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor
 C:Genetics:
 A:Gene: VCA0250
 A:Map position: 2
 C:Superfamily: mammalian alpha-amylase; alpha-amylase core homology

Query Match 75.6%; Score 31; DB 2; Length 466;
 Best Local Similarity 83.3%; Pred. No. 12e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIVLC 6
 ||:||||
 DB 4 SYIVLC 9

Search completed: August 16, 2001, 11:41:44
 Job time: 167 sec

